



OY	1088	GAGCACGAGGAGAGCTGACGAGCAGACAGACTTCTGTGGGGAGGGAGCCCTG---	11444
Db	524	GUUThrGluGlnGluLeuAlaArgLInHisSgLn--AlaLeuMetAspRPrProTyGlyLeuSpr	542
OY	1145	ACCAATGCCCGGGAGGGGCTCCACAGAAAGTGAGAGACAACAGAAAGACTGTGAAGAGAG	1204
Db	543	ArgLeuProGluGlnGlySgLinAlaHisLaIcGlnAlaGlyValGln---ValysSgLinGlu	561
OY	1205	CAGCAGAGAAGATGGGAGAGAGAGCAGCATGTGCATTGCACTGAGTTAAGCAGAGAGAGGC	1264
Db	562	ProIIeGlnSerAspArgGlnGlnGlnLaGlu-----ProProlArgGlnVal	576
OY	1265	GAGAGTGGTGTGAGAGAGGGGGGGCCGCAACTTGGAGAGACCCTGTGTGATACAAAATAATG	1324
Db	577	GluProGluGlnArgGln--ProSerGlnGlnGlnLeuLeuPheArgGlnIndLaLeu	595
OY	1335	TTCCTCAGATGCCCAACGCGCTGCAACCTTTGACAGTGTATCCAAAGCGCCCTGACCGCTGGCC	1384
Db	596	LeuLeuGlnGlnGlnArgGlnLInHisSgLinLeuAlaArgSntyGlnAlaSerMetClnAla	615
OY	1385	ACTGTGCC-----CACCAAAGCCTGGGCGCGTAACCAATCTCCCTCTCT	1429
Db	616	GlyIleProValSerPheGlnGlyGlnHisArgProLeuSerArgAlaGlnSerProAla	635
OY	1430	GCC-----CCTGGGGGATGAGAACCCCCCGACACCAACCGGTGAAGCAACTCTTC	1480
Db	636	SerAlaThrPheProValSerValGlnGlnProThrLysTrp-----ArgPhe	652
OY	1481	ACCAACAACTGTGTCTACGACAGCGTTCTATGCTAAAGCACCAGTGCATGTGCGGGAAACA	1540
Db	653	ThrThrGlnLeuValTyrAspThrLeuMetLeuLynHisGlnCyshnCysGlySerSer	672
OY	1541	CACGTGCACCCCTGACATCTGTCGGCCGAGTCSAGACATGTGGTCCCGGCTGCAGAGACA	1600
Db	673	SerSerHisProGlnLInHisAlaGlyArgLInGlnSerLLeTPserArgPheGlnIndLThr	692
OY	1601	GGCCGCGCTTAGCAGTGCAGAGCGGATCCGAGTCCGGAAGCCACAGCTGATGATGATCCAG	1660
Db	693	GlyLeuAlaGlnGlyLysCysGlnCysLLeaGlnGlyArgLysAlaThrLeuIndLcLLeuGln	712
OY	1661	ACAGTGCACCTGAAATCCACACCCTGTCTATAGGAGACCAAGTCCCCTCAACCGGACAGAG	1720
Db	713	ThrValHisSerGlnAlaHisLInThrLeuLtyGlyThrAsnProLeuAsnArgGlnLys	732
OY	1721	CTACACACAGCAAGATTGCTCCGCTCCCATGACCGACAAATGTANTGTCTGTCTCTTGT	1780
Db	733	LeuAspSerLysLysLeuLeuGlySerLeuAla--SerValPheValArgPheProCys	751
OY	1781	GGGGGCATCGGGGGGAGCAGAGGACGACCGGTGTGAATGAGATGACATCCCTCCACGTCTGTG	1840
Db	752	GlyGlyValGlyValAlaIspSerAspThrLLeTPrasGlnValHisSerLlaGlyAla	771
OY	1841	CGCATGCGAGTGGGCTGCTCTGTGAGACTGGCTCTTCAAGTGGCTGCAGAGAGCTCAAG	1900
Db	772	ArgLeuAlaValGlyLysValAlaGlnLeuValPheLtyValAlaIthrGlyLtyLeuLys	791
OY	1901	AATGAGATTTCGATCATTCGCGGCCCGGACAGACACACCGCGAGAAATCCAAACCATGGGA	1960
Db	792	AsnGlyPheAlaValAlaIspProProGlnHisLInHisAlaGlnLtySerThrPrometGly	811
OY	1961	TTCGTGCTTCACACCTGTGAGGCATCAACCGGAAACCTCTACAGACAGAGTTGAATG	2020
Db	812	PheCysTyrTrpHisnSerValAlaValAlaValLysLeuGlnGlnIthrGlyLeuSerVal	831
OY	2021	GGCAAGTCCSCATCCSGAGATGGGAGCATTCACATTCAGATGAGACACGAGAGCGGTTC	2080
Db	832	SerLysLLeuLInLeuAlaIspTrpAspValAlaHisLysGlyAsnGlyThrGlnGlnAlaPhe	851
OY	2081	TACATACACCCCTGTGTCTCTACATCTCTCTGATCCGCTATGACAGACGGGAATCTTCT	2140
Db	852	TyrSerAspProSerValLeuLtyGlyMetSerLeuHisLInArgTyrAspAspGlyAsnPheHe	871
OY	2141	CSAAGCTCTGGGGCTCTGAGAAGATTGGTGTGAAGACACAGCGCTGGGGGTACATGTGTAC	2200

[illegible]

Best Local Similarity: 59.00% Mismatches: 196  
 Query Match: 50.22% Indels: 81  
 DB: 10 Gaps: 23  
 US-09-502-945-2 (1-2885) x US-09-817-913-7 (1-967)

QY 2 GAATTCCTGTGTCGAAGTCAAGAGCCACACAGCGCGCTCAACATTCCTCCCA 61  
 Db 60 GlnbValLeuValnLysLysAlaLeuAlaHisProAsnLeuAsnHisCylLseSer 79  
 QY 62 CAGACCCCAATCTGG-----GGAGCCACCATGCTTTTGGACAGAGTTCCT 115  
 Db 80 SerCysProArgTyrTyrGlyLysThrGlnHisSerSerLeuAspGlnSerSerPro 99  
 QY 116 CCCCAGAGCGGCCCCCTGGAGCCCTCCCTCCACAACTGCTTGGCCCTGGCCCTAC 175  
 Db 100 ProGlnSer-----GlyValSerThrSerTyrAsnHisProValLeuGlyMetYr 116  
 QY 176 GACAGTCGAGACGACTTCCCTCCGAAAACAGCTTGACACCACTTGAAGTGGCT 235  
 Db 117 AspaLalysAspaRheProLeuArgLysThrAlaSerGluProAsnLeuLysLeuArg 136  
 QY 236 TCAAGGCTAAACAGAGGTGGCTGAGCGGAGAGAGTCCCTCTGCTGCGAGAGAT 295  
 Db 137 SerArgLeuLysGlnLysValAlaGluArgArgSerSerProLeuLeuArgArgLysAsp 156  
 QY 296 GGCAGCTTATAGACCTTTAAGAGAGAGCTTGAGATCAGAGTGGCGGCTGGG 355  
 Db 157 GlyProValAlaThrAlaLeuLysLysArgProLeuAspValThr----- 171  
 QY 356 GCGTCGCTGCTGTAAACAGCGACCGCGCTCCGCGCCAGCTCTCC---AACAGTCC 412  
 Db 172 ---AsperAlaCysSerSerAlaProGlySerGlyProSerSerProAsnHisSer 190  
 QY 413 CACAGCAGCACTGCTGAATGCTTACTGCTAGTCCCAACATCCCATGAGATG 472  
 Db 191 GlySerAlaSerAlaGluAsnGlyIleAlaProAlaValProSerIleProAlaGluThr 210  
 QY 473 CTCCTGAGACGAGCGCTCCCTGAGCAGCTCCCAACAGCTTACAGCTTACAGC 532  
 Db 211 SerLeuAlaHisArgLeuValAlaArgGlnGlySerAlaAlaProLeuProLeuTyrThr 230  
 QY 533 TCTCTCTCTGCGCAACATCTCCCTGAGGCTGAGGCGACAGTCACTGTCACCACTCA 592  
 Db 231 SerProSerLeuProAsnIleThrLeuGlyLeuProAlaThr----- 244  
 QY 593 CACCTACTGCTCCCGGAAGTGTGCACA-----CAGCAGAGGCGCGAGAGCGAGCC 646  
 Db 245 -----GlyProSerAlaGlyThrAlaGlyGlnGlnAspThrGluArgLeuThr 260  
 QY 647 CTCAGTCCCTGGCGGAG-----GCTGGACAGCTGACCGGCACTTGATGATG 691  
 Db 261 LeuProAlaLeuGlnGlnAlaArgLeuSerLeuPheProGlyThrHisLeuThrProTyrLeu 280  
 QY 692 AGCAGCTCTATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 751  
 Db 281 SerThrSer-----ProLeuGlnAlaGlyAspGlyGlyAla 291  
 QY 752 CAGGCGCATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 811  
 Db 292 ---AlaHisSerProLeuGlnHisMetValLeuLeuGlnGlnProProAlaGlnAla 310  
 QY 812 ACCCTACTT-----GCTGTCACCTGACAGGAGTCCCACTAGTACGGGT 859  
 Db 311 ProLeuValThrGlyLeuGlyAlaLeuProLeuHisAlaGlnSer---LeuValGlyAla 329  
 QY 860 GAAGTGTGGCCACACATGCGAGCGTAGAGGAGAGTCCCGCGAGTCCGCGCCCTAGC 919  
 Db 330 AspaArgAlaSerProSer-----IleHisLysLeuArgGlnHisArgProLeuGly 346  
 QY 920 CGCAGCTAGTCTCAGCGCTGCGCGAGAGTCCCGAGGCGCTGACAGAGTGGTCAAG 979  
 Db 347 ArgThrGlnSerAlaProLeuProGlnAsnAlaGlnAlaLeuGlnHisLeuValIleGln 366

QY 980 CACAGCAGCAGCAGTCTCTGAGAGCAGAG-----CAGCAGCAGTACAG 1027  
 Db 367 GlnGlnHisGlnGlnPheLeuGlnLysHisLysGlnGlnPheGlnGlnGlnGln 386  
 QY 1028 CTGGGCAAGATCTCTACCAACAGCAGGAGAGTGGCCAGGCGCCACACCCCTGAG 1087  
 Db 387 MetAsnLysIleIleLeuProLysProSerGluProAlaArgGlnProGlnSerHisProGlu 406  
 QY 1088 GAGCAGAGAGAGAGCTGACGCGAGCAGCAGAGTCTTGTGGGGGAGGAGCCCTG--- 1144  
 Db 407 GlnThrGlnGlnGlnLysLeuArgGlnHisGln---AlaLeuLeuAspGluProTyrLeuAsp 425  
 QY 1145 ACCATGCCCGGAGGGGCTCCACAGAGGTGAGAGCAGCAGAGAGAGCTGGAGAGAG 1204  
 Db 426 ArgLeuProGlyGlnLysGlnAlaHisAlaGlnAlaGlnValGln---ValLysGlnGlu 444  
 QY 1205 GACGAGAGAGAGATGGGAGGAGAGAGAGATGCACTCCAGTAAAGAGAGAGAGCC 1264  
 Db 445 ProIleGlnSerAspGlnGlnGlnAlaGlu-----ProProArgGlnVal 459  
 QY 1265 GAGAGTGGTCTGAGAGAGGCGCCGACTTGAGAGAGCTGCTGATACAAAAGT 1324  
 Db 460 GlnProGlnGlnArgGln---ProSerGlnGlnGlnLeuLeuPheArgGlnGlnAlaLeu 478  
 QY 1325 TTCTCAGATGCCCAACCGCTGCAACCTTTGAGAGTATACAGGCGCCCTCAGCTGGCC 1384  
 Db 479 LeuLeuGlnGlnGlnArgIleHisGlnLeuArgAsnTyrGlnAlaSerMetGlnAlaAla 498  
 QY 1385 ACTGTGCC-----CACCAAGCCCTGGGCGCTTACCAATCTCCCTGCT 1429  
 Db 499 GlyLeuProValSerPheGlyGlnHisArgProLeuSerArgAlaGlnSerSerProAla 518  
 QY 1430 GCC-----CCTGGGCGCATGAGAGAGCCCAAGACCAACCCGTAAAGACCTCTTC 1480  
 Db 519 SerAlaThrPheProValSerValGlnGlnProProThrLysPro-----ArgPhe 535  
 QY 1481 ACCAGAGTGGCTTACAGACAGCTGATGAAGACAGCAGTGCATGCTGAGGAGAGACA 1540  
 Db 536 ThrThrGlnLeuValIlyTyrAspThrLeuMetLeuLysHisGlnCysThrCysGlySerSer 555  
 QY 1541 CACGTGACCCCTGAGCAGTGTGGCCGAGATCCAGAGCATCTGTGCTCCGCTGACAGAGA 1600  
 Db 556 SerSerHisProGlnHisAlaGlnAlaArgIleGlnSerIleThrSerArgLeuGlnGluThr 575  
 QY 1601 GGCCTGCTTACAGAGTCCGAGCGGATCCGAGAGTGGCAAAGCCATGATGAGATCCAG 1660  
 Db 576 GlyLeuArgGlyLysCysGlnCysIleArgGlyArgLysAlaThrLeuGlnGlnLeuGln 595  
 QY 1661 ACAGTGCAGCTGAAATCCACACCTGCTGATGAGACAGTCCCTCAACCGGCGAGAG 1720  
 Db 596 ThrValHisSerGlnAlaHisThrLeuLeuValGlyThrAsnProLeuAsnAlaGlnLys 615  
 QY 1721 CTAGACAGCAAGAGTGTGCTGCTCCATCAGCCAGAGAGATGATGCTGTGCTGCTGCT 1780  
 Db 616 LeuAspSerLysLysLeuLeuGlnLysSerLeuAla---SerAlaPheValArgLeuProCys 634  
 QY 1781 GGGGCGATCGGGGTGAGACAGTACACCGTGTGATGATGATGATGATGATGATGATG 1840  
 Db 635 GlyGlyAlaGlyAlaAspSerAspThrIleTyrAsnGlnValHisSerAlaGlyAlaAla 654  
 QY 1841 CGCATGCGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1900  
 Db 655 ArgLeuAlaValGlyCysValAlaGlnLeuValPheLysValAlaThrGlnGlyLeuLys 674  
 QY 1901 AATGATTTCCATCATCCGCGCCCGAGAGACACACCGCAGAGATCCACAGGCATGGGA 1960  
 Db 675 AsnGlyPheAlaValAlaArgProProGlnHisHisAlaGlnGlnSerThrPrometGly 694  
 QY 1961 TTCTGCTTTTCAACTGTGAGCCATACCGCAAACTCTTACAGCAGAACTGAACTG 2020  
 Db 695 PheCysTyrPheAsnSerValAlaValAlaLysLeuLeuGlnGlnArgLeuSerVal 714

```

OY 2021 GGCAGAGTCTCATCTGACACTGGACATTCACCATGGCAATGGCCACCGAGCGCTTC 2080
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 715 SerLysIleLeuValLeuValSerPaspValHisHisGlyAsnIleGlyThrGlnGlnAlaPhe 734
OY 2081 TACAAATGACCCCTCTGCTGCTACATCTCTGATCGGATGACAGCGGAATCTCTT 2140
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 735 TyrSerAspProSerValLeuTyrMetSerLeuHisArgTyrAspSpGlyAsnPhePhe 754
OY 2141 CCAAGCTCTGGGGCTCTCCAGAGAGTGGTGGAGAGCAAGCGCTGGGGTACATGTGAAC 2200
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 755 ProGlySerGlyAlaProAspGluValGlyThrGlyProGlyValGlyIleGlyHisValAsn 774
OY 2201 GTCGCACTGGACAGAGAGTGTGGACCCCCCATTTGGAGAGCTGGAGTACCTTACAGCTTC 2260
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 775 MetAlaPheThrGlyIleGlyLeuAspProPheMetGlyAspAlaGluTyrLeuAlaAlaPhe 794
OY 2261 AGCAGAGTGGTATGCCCTTGGCCCATGGCCAGAGTTCTCACCTGATGTGGTCTTACGTCCGCC 2320
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 795 ArgThrValValMetProIleAlaSerGluPheAlaProAspValValLeuAlaSerSer 814
OY 2321 GGGTTTGATGCTGTGGAAAGACATCTGCTCTCTGGTGGCTACTCTGTCAACCGCCAGA 2380
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 815 GlyPheAspAlaValGluGlyHisProThrProLeuGlyGlyTyrAsnLeuSerAlaArg 834
OY 2381 TGTGTTGGGCACTTGACACAGACGTGATGACCTGGCAGAGGGGGCGGGTGGCTGGCC 2440
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 835 CysPheGlyTyrLeuThrLysGlnLeuMetGlyLeuAlaGlyIleArgIleValLeuAla 854
OY 2441 CTGAGAGGAGGCGCATCTTGACCGCCCATCTGATGATCGCTGGAACCTGTGTCTGGCT 2500
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 855 LeuGluGlyGlyHisAspLeuThrAlaIleCysAspAlaSerGluAlaGlyValSerAla 874
OY 2501 CTGCTCAGTGTAAAGCTGCAGCCCTTGATGAGGCACTTTGACAGAAAGCCCAACATC 2560
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 875 LeuLeuGlyAsnGlyIleuLeuAspProLeuProGlyIleValLeuGlnGlnArgProAsnAla 894
OY 2561 AACGCGCGCCAGCTAGACAAAGTCATCGAGATCCAGAAACAGCACTGAGCTGTGG 2620
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 895 AsnAlaValArgSerMetCyluysValMetGluIleHisSerLysTyrTrpArgCysLeu 914
OY 2621 CAGAAATTCGCCCTGCTGTCGGCCGCTCCCTGCGAGGGGCCCAAGCAGATGAGACCGAA 2680
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 915 GlnArgThrThrSerThrAlaGlyArgSerLeuIleGluAlaGlnThrCysGlnAsnGlu 934
OY 2681 GAAGCCGAAT-CTGAACGCATATGCGCTTGTGCTGGGGGGCCGACAGCGCCCAAGCT 2739
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 935 GluAlaGluThrValThrAlaMetAlaSerLeuSerValGlyValLysProAlaGlu--- 953
OY 2740 GCGGCAAGCCCGGACACAGCCCGGCGGAGAGAGAGCCCATGGAGAGAGCTGCC 2799
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 954 -----LysArgProAspGluGluProMetGluGluProPro 966
OY 2800 CTG 2802
    |||
Db 967 Leu 967

```

```

; SEQ ID NO 7
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Human
US-09-817-538-7

Alignment Scores:
Pred. No.: 1.6e-124 Length: 967
Score: 2685.00 Matches: 567
Percent Similarity: 71.28% Conservative: 118
Best Local Similarity: 59.00% Mismatches: 196
Query Match: 50.22% Indels: 81
DB: 10 Gaps: 23

US-09-502-945-2 (1-2885) x US-09-817-538-7 (1-967)
OY 2 GAATTCCTGTGCGAAGTCAAGAGAGCCACACAGCGGGGCTCAACCATTCCTCCCA 61
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 60 GluPheValLeuAsnLysLysAlaLeuAlaHisProAsnLeuAsnHisCysIleSer 79
OY 62 CAGCAGCCCAAAATGCTGG-----GGAGCCACATGCTTTTGGACAGAGTTCCCT 115
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 80 SerCysProArgTyrTrpTyrGlyLysThrGlnHisSerSerLeuAspGlnSerPro 99
OY 116 CCCCAGAGCGGGCCCCCTGGAGCCCTCCCTCTACAACTGCTTTGGCTGGCCCTAC 175
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 100 ProGlnSer-----GlyValSerThrSerTyrAsnHisProValLeuGlyMetTyr 116
OY 176 GACAGTCGAGAGACTTCCCTCCGCAAAACAGCCTTGAAACCCCAACTTGAAGTGCCT 235
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 117 AspAlaIleAspAspPheProLeuArgTyrThrAlaSerGluProAsnLeuLysLeuArg 136
OY 236 TCAAGGCTAAACAGAAAGTGGCTGAGCGAGAGACAGTCCCTCGTGGTGCAGAGAT 295
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 137 SerArgLeuLysGlnLysValAlaGluArgArgSerSerProLeuLeuArgTyrAsp 156
OY 296 GGGACTGTATATGACACCTTTAAGAAAGACGCTGTTGAGATACAGAGTCCCGGCTGGG 355
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 157 GlyProValValThrAlaLeuLysLysArgProLeuAspValThr----- 171
OY 356 GCGTGCCTGCTGTAAACAGCGCACCGCGGCTCGGCGCCAGCTCTCC--AACAGCTCC 412
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 172 ---AspSerAlaCysSerSerAlaProGlySerGlyProSerSerProAsnAsnSerSer 190
OY 413 CACAGCACCATGCTGAGATGCTTACTGCTGAGTCCAGTCCCAACATGCCCACTGAGATG 472
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 191 GlySerValSerAlaGluAsnGlyIleAlaProAlaValProSerLleProAlaGluThr 210
OY 473 CTCCCTCAGACACCGACCCCTCCCTGAGACAGCTGCCCAACACAGTTACAGCTTACAGG 532
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 211 SerLeuAlaHisArgLeuValAlaArgGluGlySerAlaAlaProLeuProLeuTyrThr 230
OY 533 TCTCTCTCTGCGCAACATCTCCCTAGGGCTGACGGCCAGCGTCACTGCAACATCA 592
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 231 SerProSerLeuProAsnLleThrLeuGlyLeuProAlaThr----- 244
OY 593 CACTCAGTGCCTCCCGAAGCTGTGCACA-----CAGCAGAGAGCCGAGAGCGAGGCC 646
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 245 -----GlyProSerAlaGlyThrAlaGlyGlnGlnAspThrClnArgLeuThr 260
OY 647 CTCAGATCCCTGCGGACAG-----GGTGGACAGCTTACCGGCAAGTTTCATG 691
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 261 LeuProAlaLeuGlnGlnArgLeuSerLeuPheProGlyThrHisLeuThrProTyrLeu 280
OY 692 AGCAGATCTCTATTCCTGCGTGGCTGCTGGGCGGAGCACTGGAGGGGAGCGGAGCCCC 751
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 281 SerThrSer-----ProLeuGlnArgAspGlyAla 291
OY 752 CACGGGCAATGCTCCCTGTCAGAGCATGTGCTGTGAGACAGAGCCGCGGACAGAGAGC 811
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 292 ---AlaHisSerProLeuLeuGlnHisMetValLeuLeuGluGlnProProAlaGlnAla 310
OY 812 ACCCTCATTT-----GCTGTGCCACTTCCAGGCGGAGTCCCACTAGTACGGGT 859

```



Db 311 ProLeuValThrGlyLeuGlyAlaLeuProLeuHisAlaGlnSer---LeuValGlyAla 329  
|||:::||||| ||||| |||||  
QY 860 GAACSTGTGGCCACACGATCGCGAGCGTAGACACSTCCCGCGCATGGCCCTGACG 919  
|||::: |||::: |||||::: |||||  
Db 330 AspArgValSerProSer-----IleHisGlyLeuAlaGlnHisAlaArgProLeuGly 346  
QY 920 CGACATGATCTCCATCGCGTGGCGGACAGTCCCGACGCGCTGACAGATGGATGACGA 979  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 347 ArgThrGlnSerAlaProLeuProGlnAsnAlaGlnAlaLeuGlnHisLeuValIleGln 366  
QY 980 CAACAGCACACAGATCTCTGTGAGAAAGCAAG-----CACAGACAGTACAG 1027  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 367 GlnGlnHisGlnGlnIleLeuGlnIuLysHisGlnGlnIlePheGlnGlnGlnIleGln 386  
QY 1028 CTGGGCAAGATCTCTACACAGACAGGAGAGCTCCAGCAGCAGCCACACCTGTAG 1087  
::: |||||::: ||| |||||::: |||||::: |||||  
Db 387 MetAsnIuLysIleIleLeuProLysProSerGluProAlaAlaArgGlnProGlnSerHisProGln 406  
QY 1088 GAGACGAGAGAGAGCTGACGGGAGCGAGAGCTCTCTGGGGGAGGAGCCCTG--- 1144  
|||||::: |||||::: ||| ||||| ||||| |||||  
Db 407 GluThrGlnGlnGlnIleuLysArgGlnHisGln---AlaLeuLeuAspGluProGluTyrLeuAsp 425  
QY 1145 ACCATGCCCCGGAGGGGCTCCACAGAGATGAGACACACAGAGAACCTGAGAGAGAG 1204  
::: |||::: |||::: |||::: |||::: |||  
Db 426 ArgLeuProGluGlnIuLysGlnAlaHisAlaGlnAlaGlyValGln---ValLysGlnGln 444  
QY 1205 GACGAGGAAGAGATGGGAGAGGAGGAGAGATGCGATTCAGGTTAAGACAGAGAGGCG 1264  
||| ||| ||||| |||||  
Db 445 ProIleGlnSerAspGluIuGlnAlaGln-----ProProArgGluVal 459  
QY 1265 GAGAGTGTGTGAGAGAGGAGGCGGCTGAGAGAGCTGTGCTGGATACAAAATCTG 1324  
||| |||::: |||::: |||::: |||::: |||  
Db 460 GluProGluGlnAlaArgGln---ProSerGlnGlnGlnIuLeuLeuPheArgGlnGlnAlaLeu 478  
QY 1325 TTCTGACATGCCCAACCGCTGCAACCTTTGCAAGGTATCCAAAGCGCCCTGACGCTGGCC 1384  
::: |||::: |||::: |||::: |||::: |||  
Db 479 LeuLeuGlnGlnGlnIleArgGlnIleHisGlnLeuArgAsnTyrGlnAlaSerMetGluAlaAla 498  
QY 1385 ACTGTGCCC-----CACCAAGCCCTGGGCGGTACCCAAATCCCTCCCTGCT 1429  
|||::: |||::: ||| ||| ||||| ||||| |||||  
Db 499 GlyIleIleProValSerPheGlyLeuGlyHisArgProLeuSerArgAlaIleSerSerProAla 518  
QY 1430 GCC-----CTGGGCGGATGAAGAGACCCCGACACCAACCGCTCAAGCAACCTCTTC 1480  
::: |||::: |||::: |||::: |||::: |||  
Db 519 SerAlaThrPheProValSerValGlnGluProProThrLysPro-----ArgPhe 535  
QY 1481 ACCCAAGATGTGTCTACACAGACGTTGATTAAGACACACATGATGTGCGGGAACACA 1540  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 536 ThrThrGlyLeuValTyrAspThrLeuMetLeuLysGlnSerHisGlnSerHisGlySerSer 555  
QY 1541 CACGTGCAACCGTACGATCTGGCGGATCCAGACATGTGGTCCGCGCTGACAGAGACA 1600  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 556 SerSerHisProGlnHisIleArgIleGlnSerIleTyrSerIleGlnGlnIuThr 575  
QY 1601 GGCCTGCTTAGCAAGTGCAGGCGATCCAGGTCCAAAGCCACGCTAATGATGATCAG 1660  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 576 GlyLeuArgGlyLysCysGlnCysIleArgGlyArgLysAlaThrLeuGlnIuLeuGln 595  
QY 1661 ACAAGTGCATCTGATACACACACCTGTCTATGGACAGATCCCTCAACCGGACGAAG 1720  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 596 ThrValHisSerGlnValHisThrLeuLeuTyrGlyThrAsnProLeuAsnArgGlnLys 615  
QY 1721 CTAGACAGCAAGATGTGCTGCCATCAGCGACAAATGATGCTGTGCTGCTTGT 1780  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 616 LeuAspSerLysLysLeuLeuGlnGlySerLeuAla---SerValPheValArgLeuProCys 634  
QY 1781 GGGGCGATCGGGGTGACATGACACACCTGTGATGAGATGACATGCACATCTGCTGTG 1840  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 635 GlyGlyValGlyValAspSerAspThrIleTyrAsnGlnValHisSerIleAlaAla 654  
QY 1841 CGCATGCGAGTGGGCTGCTGCTGAGAGCTGCGCTTCAAGGTGGCTGACGAGAGCTCAAG 1900  
|||||::: |||||::: |||||::: |||||::: |||||

Db 655 ArgLeuAlaValGlyCysValValGlyLeuLeuValPheLeuValAlaThrGlnGlyLeuLys 674  
QY 1901 AATGATTTGGCATCATCCGGGCCCCAGACACACACCGGAGAAATCCACAGCATGGGA 1960  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 675 AsnGlyPheAlaValAlaArgProProGlnHisHisAlaGlnIuLeuSerThrPrometGly 694  
QY 1961 TTCTGCTCTCAACTGTGATGACATCCACCGGAAATCTCTACAGAGAAATGTAACG 2020  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 695 PheCysTyrPheAsnSerValAlaValAlaAlaLysLeuLeuGlnIuThrGlyLeuSerVal 714  
QY 2021 GGCAGAGCTCTCATCTGTGACATGGACATTCACATGGACATGGACATGGACATGGAC 2080  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 715 SerLysIleLeuIleValAspTyrAspValHisHisGlyAsnGlyThrGlnGlnAlaPhe 734  
QY 2081 TACATGACCCCTGTGTCTTACATCTGTGTGATGCTGTGATGACACAGGAACTTCTTT 2140  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 735 TyrSerAspProSerValIleuTyrMetSerLeuHisArgTyrAspAspGlyAsnPhePhe 754  
QY 2141 CCAGGCTCTGGGCTCTGTAAGAGGTTGGTGGAGGACAGAGGCGTGGGATACATGTGAAC 2200  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 755 ProGlySerGlyAlaProAspGluValGlyThrGlyProGlyValAlaLysPheAsnValAsn 774  
QY 2201 GTGGCATGACAGAGAGTGTGACACCCCGCATTTGAGACGTGAGATCTTACAGCCTTC 2260  
::: |||::: |||::: |||::: |||::: |||  
Db 775 MetAlaPheThrGlyLeuLeuAspProprometGlyAspAlaGlnIuTyrLeuAlaAlaPhe 794  
QY 2261 AGACACGTGTGATGCGCATTTGCGCCACAGATTCTACCTGATGTGGTCTAGTCTCGCC 2320  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 795 ArgThrValValMetProIleAlaSerGluPheAlaProAspValValLeuAlaSerSer 814  
QY 2321 GGGTTGATGCTGTGTAAGACATCTGTCTGCTGGTGGTGGTGTCTGACACCGCCAGA 2380  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 815 GlyPheAspAlaValAlaGlnIuLysHisProThrProLeuGlyGlyTyrHisLeuSerAlaArg 834  
QY 2381 TGTTTTGGCCACTTGCACACAGACGTGATGACCTGTGACAGGAGGCGCGGTGTGCTGGCC 2440  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 835 CysPheGlyTyrLeuThrLysGlnLeuMetGlyLeuAlaGlyGlyArgIleValIleuAla 854  
QY 2441 CTGGAGGAGGCGCATGTGACACCGCATCTGTGATGATGCTGTGACCTGTGTCTGGCT 2500  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 855 LeuGlnGlyLeuHisAspLeuThrAlaIleCysAspAlaSerIuLysValSerAla 874  
QY 2501 CTGCTCAGTGTAAAGCTGTGACGCGCTGTGATGAGCGAGCTGTTCGACAGAAAGCCCAATC 2560  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 875 LeuLeuGlyAsnGluLeuAspProLeuProGluLysValLeuGlnIuThrProAsnAla 894  
QY 2561 AACGCACTGGCCACGCTTAAGAAATGATGACATGACATCCAGACAAACATCTGAGCTGTG 2620  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 895 AsnAlaValAlaArgSerMetGluLysValIleMetGluIleHisSerLysTyrTrpArgCysLeu 914  
QY 2621 CAGAAATTTGGCGCTGTGCGCGCGTCCCTGCGAGGGGCCCCAACAGATGAGACCGAA 2680  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 915 GlnArgThrThrSerThrAlaGlyArgSerLeuIleGlnAlaGlnThrCysGlnAsnGln 934  
QY 2681 GAAGCCGAAT-GTGAACCCCATGCGCTGTGTTGGTGGGGCCGACAGGCGCAAGCT 2739  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 935 GluAlaGlnIuThrValThrAlaMetAlaSerLeuSerValGlyValLysProAlaGln--- 953  
QY 2740 GCGGCAAGCCCGGAGACACAGCCCGAGGCGGAGAGAGCCCATGAGACAGAGCTGCG 2799  
|||||::: |||||::: |||||::: |||||::: |||||  
Db 954 -----LysArgProAspGlnGluIuPrometGlnIuGlnIuProPro 966  
QY 2800 CTG 2802  
|||  
Db 967 Leu 967

RESULT 8  
US-10-072-094-87  
; Sequence 87, Application US/10072094  
; Publication No. US20030025538A1  
; GENERAL INFORMATION:  
; APPLICANT: JACKSON, DONALD  
; APPLICANT: LORENZI, MATTHEW

```
; APPLICANT: ATTAR, RICARDO
; APPLICANT: GOTTARDIS, MARCO
; TITLE OF INVENTION: NOVEL HUMAN HISTONE PEPTIDES
; FILE REFERENCE: 3053-414US1
; CURRENT APPLICATION NUMBER: US/10/072.094
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/296,296
; PRIORITY FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 87
; LENGTH: 1069
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-072-094-87

Alignment Scores:
Pred. NO.: 1.09e-123 Length: 1069
Score: 2668.00 Matches: 570
Percent Similarity: 70.88% Conservative: 114
Best Local Similarity: 59.07% Mismatches: 203
Query Match: 49.91% Indels: 79
DB: 9 Gaps: 21

US-09-502-945-2 (1-2885) x US-10-072-094-87 (1-1069)
QY 2 GAATTCCTCTGTGCAAGTCA-----AAGAGCCACACAGCGCGCTCAACCATTC 55
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 152 GluPheLeuLeuSerLysSerLalaThrLysAspThrProHnAsnGlyLysAsnHisSer 171
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 56 CTCGCCAGACCCCAATGCTG-----GGAGCCACATGCTTCTTGGACAGAGT 109
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Db 172 ValSerArgHisProLysLeuThrPyrThrAlaAlaHisHisThrSerLeuAspLnsr 191
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 110 TCCCTCCCGAGAGCGCGCCCTGGAGCGCTCCCTCTCAAACTGCTTGGCTGG 169
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 192 SerProPheLeuSer-----GlyThrSerProSerTyrLysTyrThrLeuProGly 208
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 170 CCTTACGACAGTGCAGACACTCCCTCCGCAAAACAGCTGTGAACCCCACTTGA 229
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 209 AlaGlnAspAlaLysAspAspPheProLeuArgLysThrAlaSerGluProAsnLeuLys 228
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 230 GTGCGTTCAAGGCTTAAACAGAGGTGCTAGCGGAGACAGTCCCTCGCTGCTGC 289
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 229 ValArgSerArgLeuLysGlnLysValAlaGluArgArgSerProLeuLeuArgArg 248
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 290 AAGATGGGACTGTTATAGACCTTTAAGAAGAGCTGTGAGATACAGAGTCCCGG 349
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 249 LysAspGlyAsnValValThrSerPheLysLysArgMetPheGluValThr----- 265
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 350 CCTGGGCGGCTCGCTGTGAACAGCGACCGGCTCGGCGCCAGCTTCCCAACAGC 409
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 266 -----GluSerSerValSerSerSerProGlySerGlyProSerSerProAsn--- 282
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 410 TCCACAGACACCATCGCTGAGATGGCTTTACTAGTCCATGCCCAACATCCCACTG 469
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 283 -----AsnGlyProThrGlySerValThrGlnAsnGluThrSer 295
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 470 ATGCTCCCT-----CAGCACCGAGCCCTCCCTCTG 499
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 296 ValLeuProThrProHisAlaGluGlnMetValSerGlnGlnArgLysLeuLeuHis 315
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 500 GACAGCTCCCCACCACTGATCGCTACAGCTCTCTTCTGCGCCCAACATGCCCTA 559
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 316 GluAspSerMetAsnLeuLeuSerLeuTyrThrSerProSerLeuProAsnLethrLeu 335
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 560 GGGCTGAGCGCCAGCTCATGTCTACCAACTACACCTCATGCTCCCGCAAGCTGTGC 619
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 336 GlyLeuProAla-----ValProSerGlnLeuAsnAlaSerAsnSerLeuLys 351
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 620 ACACAGCAGAGGCCGAGAGCGAGCCCTCAAGTCCCTGCGGAGGTGGACAGCTGACC 679
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 352 GluLysGlnLysCysGlu-----ThrGlnThrLeuArgGlnGlyValProLeuPro 368
```

```
QY 680 GGCAAGTTC-----ATGACACATCCTTATCTTGGCTGCTGCTGCGGCTG 727
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 369 GlyLysTyrGlyLysSerLysProAlaSerSerSerHisPro-----HisVal 384
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 728 GCATCGAGGGCGAGGGAGGCCCGCCAGCGATCCCTCCCTGCTGACAGATGCTGTG 787
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 385 ThrLeuGlnLysProProAsnSerSerHisGlnAlaLeuGlnHisLeuLeu 404
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 788 CTGAGCAGCGCCCGCAGACAGACACCTCATCTCT-----GTGSCACTSCAGSGCAG 841
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 405 LysGlnGlnMetArgGlnGlnLysLeuLeuValAlaGlyLysValProLeuHisProGln 424
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 842 TCCCCACTAGTACGGGTGAAGCTGTGGCCACACAGATGCGGAGGTAGGCACTCCCG 901
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 425 SerProLeuAlaThrLysGlnArgLysLeuProGlyLysThrHisLysLeuPro 444
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 902 CGGATCGGCGCCCTGAGCGCCGACATGCTCTACCCGCTGCGGCGAGATGCCAGGCCCTG 961
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 445 ArgHisArgProLeuAsnArgThrGlnSerLysLeuProLeuProGlnSer-----ThrLeu 462
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 962 CAGCAGCTGGTCATGCAACACAGACACAGCATCTCTGAGAGCAGAG-----CAG 1015
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 463 AlaGlnLeuValIleGlnGlnGlnHisGlnLysGlnLysGlnLysGlnLysGln 482
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1016 CAGCAGCTACAGCTGGGCAAGATCTCACCAAGACAGGGAGCTGCGCCAGCCAGCC 1075
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 483 GlnGlnIleHisMetAsnLysLeuLeuSerLysSerLysGlnGlnLysGlnProGly 502
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1076 ACCCACTTGAAGACAGACAGAGAGCTGACGAGCAGCAGAGAGCTTCTGCGGAG 1135
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 503 SerHisLeuGlnGlnAlaGlnGlnGlnLeuGln-----GlyAsp 515
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1136 GGAGCCCTGACCATGCCCCCGGAGGGCTCCACAGAGATGAGACACAGAGAGACTG 1195
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 516 GlnAlaMetGlnGlnAspArgAlaProSerSer-----GlyAsnSerThrArgSerAspSer 534
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1196 GAGGAGGAGCAGAGAGAGAGATGGGAGGAGAGAGAGATTCATCCAGGTTAAGAC 1255
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 535 SerLysValLysAspArgThrLeuGly-----GlnValGlyAlaValLysValLysGln 552
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1256 GAGAGGGCGAGAGTGGTGTCTGAGAGAGGGCGCCCATTTGAGAGGCTGCTGATAC 1315
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 553 GluProValAspSer-----AspGlnAspAlaGlnIleGlnIleMetGluSerGlyGlu 570
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1316 AAAAAGCTTCTCAGATGCCCCAACCGCTGCACACT-----TTGCAGGTG 1360
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 571 GlnAlaAlaPheMetGlnGlnProPheLeuGluProThrHisThrArgAlaLeuSerVal 590
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1361 TACCAAGCGCCCTCAGCTGGCCACTGTG-----CCCAACCAAGCCCTGGGC 1408
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 591 ArgGlnAlaProLeuAlaAlaValGlyMetAspGlyLeuGlnLysHisArgLeuValSer 610
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1409 CGTACCCCAATCTCCCTGCTGCTGCGCCCTGGGGGCAATGAAGAAACCCCAAGACCAACCCGTC 1468
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 611 ArgThrHisSerSerProAlaAlaSerValLeuProHisProAlaMetAspArgProLeu 630
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1469 AAGCACTCTTACCAACAGAGTGTGCTACAGACAGTTCATGCTAAAGCAACATGTGATG 1528
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 631 GlnProGlySerAlaThrGlyIleAlaLysTyrAspProLeuMetLysHisGlnCysVal 650
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1529 TGCAGAAACACACAGCTGACATGCTGAGCCGCGATCCAGACATCTGCTCCCG 1588
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 651 CysGlnLysAsnSerThrThrHisProGlnHisAlaGlyArgLysIleThrPserArg 670
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1589 CTGAGAGAGACAGGCTGCTTAGCAAGTGCAGCGAGTCCAGAGTGCCTCAAGCCAGCTA 1648
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 671 LeuGlnGlnThrGlyLeuLeuAsnLysCysGlnArgLysGlnGlnArgLysAlaSerLeu 690
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1649 GATAGATCCAGACAGTGCATGTGATACACACCTGCTCTTGGGACACAGTCCCTC 1708
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 691 GlnIleIleGlnLeuValHisSerGlnHisHisSerLeuLeuTyrGlyThrAsnProLeu 710
```

```
QY 1709 AACGGCAGAGCTAGACAGAGAGTTGCTGCTCCATCCAGACGAGATGATGCT 1768
    ::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 711 AspGlyGlnIleuAspProArgIleIleuGlyAspSerGlnIleuPhePheSer 730
QY 1769 GTGTCCTGTTGGGGGCGATGGGGGTGACAGTGAACCGCTGTGAATGATGACATCC 1828
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 731 SerIleuProCysGlyGlyIleuGlyIleuAspSerAspThrIleuPheGlnIleuHisSer 750
QY 1829 TCCAGGCTGTGGAGAGTGGAGTGGGCTGCTGAGTGGGCTTCAAGTGCGGTGCA 1888
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 751 SerGlyAlaIleuAlaIleuGlyCysValIleuIleuAlaSerValAlaSer 770
QY 1889 GGAGAGCTCAAGATGATTTGCCATATCCGGCCCCAGACAGCCAGCGGAGATCC 1948
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 771 GlyGlnIleuLysGlnIleuAlaValAlaArgProProGlyIleuHisAlaGlnIleuSer 790
QY 1949 ACAGCCATGGGATTTGCTTTTCAACTGTGTGCGATCAACGCCAAATCTCTACAGCAG 2008
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 791 ThrAlaMetGlyPheCysPhePheAsnSerValAlaIleuThrAlaLysTyrIleuArgAsp 810
QY 2009 AAGTTGACGCTGGGAGGCTCATGCTGAGTGGAGCTGGACATTCACCATGGCAATGGCACC 2068
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 811 GlnIleuAsnIleuSerIleuIleuValIleuAlaSerPheAlaHisIleuGlnIleuThr 830
QY 2069 CAGCAGGCTTTACATGATGACCCCTGTGCTTACATCTCTGTGATGCTGTGATGACAC 2128
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 831 GlnGlnAlaPheTyrAlaAspProSerIleuIleuTyrIleuSerIleuHisArgTyrAspGln 850
QY 2129 GGGAACTTTCTTCCAGGCTCTGGGGCTCTGAAAGAGTGGTGGAGAGCCAGCGTGGGG 2188
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 851 GlyAsnPhePheProGlySerGlyAlaProAsnGlnIleuGlyThrGlyLeuGlyGlnIleu 870
QY 2189 TACATGTGACGCTGGCATGAGCAGAGAGTGTGACCCCCCTGGTGGAGAGCTGGAGTAC 2248
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 871 TyrAsnIleuAsnIleuAlaThrPheGlnIleuGlyIleuAspProProMetGlyAspValGlnTyr 890
QY 2249 CTTCACGCTTTCAGAGCAGTGTGATGCCCATTCGCCACAGATTTCTACCTGATGTGTC 2308
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 891 LeuGlnAlaPheArgThrIleuValIleuValIleuValIleuLysIleuPheAspProAspMetVal 910
QY 2309 CTAGTCTCCGCGGGTTTATGCTGTGTGAAGACATCTGTCTCTGCTGGTGGCTACTCT 2368
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 911 LeuValSerAlaGlyPheAspAlaLeuGlnIleuHisThrProProLeuGlnIleuTyrLys 930
QY 2369 GTACACCGCGATGTTTGGCCACTTGACACAGCAGCTGATGACCCCTGGAGGGGGCGG 2428
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 931 ValThrAlaLysCysPheGlyHisIleuThrLysGlnIleuMetThrIleuAlaAspGlyArg 950
QY 2429 GTGTCCTGGCCCTGAGGAGGCGCATGACTTGACCGCATCTGTGATGCTGTGAGCT 2488
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 951 ValValIleuAlaLeuGlnIleuGlyGlnIleuAspLeuThrAlaIleuCysAspAlaSerGlnAla 970
QY 2489 TGTGTCCTGGCTGTGCTGATGTAAGCTGACACCTTGATGATGAGCAGCTTGGCAGCA 2548
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 971 CysValAlaAsnAlaLeuLeuGlnIleuGlnIleuGlnIleuGlnIleuHisGln 990
QY 2549 AAGCCCATCAACAGCAGTGGCGCAGCTAGAGAAATGTCAGATCCAGACCAACAAAC 2608
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 991 SerProAsnMetLysAlaValIleuSerLeuGlnIleuIleuIleuIleuIleuIleuIleu 1010
QY 2609 TGGAGCTGTGTGCAAGATTCCGCGCTGTGCGCGGCTGGCGGAGGAGCGCCAAAGCA 2668
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1011 TrpLysSerValArgMetValAlaValAlaProArgGlyCysAlaLeuAlaGlyAlaGlnIleu 1030
QY 2669 GGTGAGACCGCAAGACCGCAAT-GTGAACGCCATGCGCTTGTGCTGTGGGGGCGCGAA 2727
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1031 -----GlnGlnIleuThrGlnThrValSerAlaLeuAlaSerLeuThrValAspValGln 1048
QY 2728 CAGGCCAAGCTGCGGAGCGCGGAGACACAGCCCGGAGCGGAGAGAGCCCATGAGG 2787
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1049 GlnProPheAla-----GlnGlnAspSerArgThrAlaGlyIleuProMetGln 1064
QY 2788 CAGGAGCTTGCCCTG 2802
```

```
Db 1065 GlnGlnProAlaLeu 1069
RESULT 9
US-10-072-094-89
; Sequence 89, Application US/10072094
; Publication NO. US20030025538A1
; GENERAL INFORMATION:
; APPLICANT: JACKSON, DONALD
; APPLICANT: LORENZI, MATTHEW
; APPLICANT: ATTAR, RICARDO
; APPLICANT: GOTTARDIS, MARCO
; TITLE OF INVENTION: NOVEL HUMAN HISTONE PEPTIDES
; FILE REFERENCE: 3053-4145US1
; CURRENT APPLICATION NUMBER: US/10/072,094
; PRIORITY FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,296
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 89
; LENGTH: 1011
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-072-094-89

Alignment Scores:
Pred. No.: 2,51e-119 Length: 1011
Score: 2579.00 Matches: 540
Percent Similarity: 72.05% Conservative: 107
Best Local Similarity: 60.13% Mismatches: 179
Query Match: 48.24% Indels: 72
DB: Gaps: 19

US-09-502-945-2 (1-2885) x US-10-072-094-89 (1-1011)
QY 2 GAATTCCTTTGTGGAAGTCA-----AAGGAGCCCAACAGCCGCGCTCAACATTCG 55
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 149 GlnPheLeuLeuSerIleuLysSerAlaIleuThrLysAspThrProThrAsnGlyLysAsnHisSer 168
QY 56 CTCCACAGCAGCCCAAAAGCTGG-----GAGGCCCATGCTTTTGGACAGAGT 109
    ::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 169 ValSerArgHisProLysLeuTyrThrAlaAlaHisThrSerLeuAspGlnSer 188
QY 110 TCCGCTCCCGAGGCGGCGCCCTGGGAGCGCTCCCTGACAACTGGCTTGGCTGGG 169
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 189 SerProLeuSer-----GlyThrSerProSerTyrLysTyrThrLeuProGly 205
QY 170 CCCTAGACAGTGGAGAGCACTTCCCTCCGCAAAACAGCCCTGACCCAACTTGAA 229
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 206 AlaGlnAlaPheLysAspAspPheProLeuArgLysThrAlaSerGlnProAsnLeuLys 225
QY 230 GTGCGTTCAAGGCTAAACAGAGGTGCTGAGCGGAGAGAGCATGCTCTGCTGCTGG 289
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 226 ValArgSerArgLeuLysGlnLysValAlaGlnArgArgSerSerProLeuLeuArgArg 245
QY 290 AAGGATGGGCTTTTATTGACACTTAAAGAGAGGAGCTGTGAGTCAACAGGTGCCGG 349
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 246 LysAspGlyAsnValValIleuSerPheLysLysArgMetPheGlnValIleu----- 262
QY 350 CCGGGGCGCTGCGCTGTGTAACAGCAGCCGCGCTCCGGCGCGCGCTCCCAACAGC 409
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 263 -----GlnSerSerValSerSerSerProGlySerGlyProSerSerProAsn--- 279
QY 410 TCCACAGCAGCATCGCTGAGATGAGCTTACTGCTGAGTCCGCCCAACTCCCAACAG 469
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 280 -----AsnGlyProThrGlnGlySerValThrGlnAsnGlnIleuThrSer 292
QY 470 ATGCTCCCT-----CAGCAGCAGCCCTCCCTCTG 499
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 293 ValIleuProThrProThrAlaGlnIleuMetValSerGlnArgGlyIleuLeuHis 312
QY 500 GACAGCTCCCAACCAAGTTCAAGCTCTTACAGCTGCTCTGCGCAACATCTCCGA 559
```

Dd	313	GTuAspSerMetAnLeuLeuSerLeuTyrThiSerProSerLeuProAnHisLeuThi	332
Qy	560	GGGCTGAGGCGCCACGGTACTGTCCAACTCAACCTCACTGCCTCCCGAAGCTTGC	619
Dd	333	GTLeuProAla-----ValProSerGlnLeuAnHisLaserAnSerLeuLys	348
Qy	620	ACACAGCGAGGCGCCGAGGAGGCGCCCTCACTCCCTGGCGAGGCTGGACGGTAC	679
Dd	349	GTuLysGlnLysCysGln-----ThiGlnThiLeuArgGlnLysValProLeuPro	365
Qy	680	GGCAAGTTC-----ATGAGCACAATTCCTATTCCTGGCTGCCTGGCGCTG	727
Dd	366	GTuGlnTyrGlnLysSerLeProAlaSerSerHisPro-----HisVal	381
Qy	728	GCATGGAGGGCCACGGGAGCCCCCAGGGCATGCCCTCCCTGTGAGCATGTGCTTGG	787
Dd	382	ThiLeuGlnGlnLysProProAnSerSerHisGlnAlaLeuLeuHisLeuLeu	401
Qy	788	CTGAGCAGGCGCCCGGACAGAGACCCCTATTGCT-----GTGCACTCCACGGGAG	841
Dd	402	LysGlnGlnMetArgGlnGlnLysLeuLeuValAlaGlyLysValProLeuHisProGln	421
Qy	842	TCCCCACTAGTACGCGGTGAACGTGTGGCCACACAGATCGGAGCGGTAGCAAGCTCCG	901
Dd	422	SerProLeuAlaThrLysGlnArgLysLeuSerProGlyLysArgGlyThiHisLysLeuPro	441
Qy	902	CGGCAATGCGCCCGGACCGGACACTCTCTCAACCGCTGGCGGAGTCCCGACGGCCGTG	961
Dd	442	ArgHisArgProLeuAsnAlaArgThiGlnSerAlaProLeuProGlnSer-----ThiLeu	459
Qy	962	CAGCAGCTGTATGTACAACAACAGACACAGCAAGTCTCTGAGAAAGCAGAG-----CAG	1013
Dd	460	AlaGlnLeuValIleGlnGlnGlnHisGlnGlnPheLeuGlnLysGlnLysGln	479
Qy	1016	CAGAGCTACAGCGTGGGCGAGATCCTCACAAAGGAGAGCTGCCGAGGCGGCAAC	1075
Dd	480	GlnGlnIleHisMetAnLysLeuLeuSerLysSerLeuLeuLysGlnProGly	499
Qy	1076	ACCCACCTGAGGAGACAGAGAGAGACTACGCGACAGAGAGGTCTTCTGGGGAG	1135
Dd	500	SerHisLeuGlnGlnValGlnGlnGlnLeuGln-----GlyAsp	512
Qy	1136	GGAGCCCTGACCATGCCCCGGGAGGCTCCACAGAGATGAGAGCACACAGAAAGCTG	1195
Dd	513	GlnAlaMetGlnLysAspArgAlaProSerSer--GlyAsnSerThrArgSerAspSer	531
Qy	1196	GAGGAGGAGGACAGGAAGAGATGGGAGGAGAGAGAGAGATGCTCCAGAGTAAAGAC	1255
Dd	532	SerAlaCysValAspAspThrLeuGly-----GlnValGlyAlaValLysValLysGln	549
Qy	1256	GAGGAGGCGGAGCTGCTGCTGAGAGAGGCGCCGCACTTGGAGAGGCTGTGCTGATAC	1315
Dd	550	GlnProValAspSer-----AspGlnAspAlaGlnIleGlnGlnMetGlnSerGln	567
Qy	1316	AAAAAATGTTCTCAGATGCCCAACCGCTGCACACG-----TTGGAGGG	1360
Dd	568	GlnAlaAlaPheMetGlnGlnProPheLeuGlnProThiHisThrArgAlaLeuSerVal	587
Qy	1361	TACCAAGCGCCCTCAGCCTGGCCACTGG-----CCCAACCAAGCCTGGGC	1408
Dd	588	ArgGlnAlaProLeuAlaAlaValAlaGlyMetAspGlyLeuGlnLysHisArgLeuValSer	607
Qy	1409	CGTACCCAAATCCTCCCTGCTGCTGCCCTGGGGGCGATGAAGAACCCCGACCAACCTGC	1468
Dd	608	ArgThrHisSerSerProAlaAlaSerValLeuProHisProAlaMetAspArgProLeu	627
Qy	1469	AAGACACCTTCAACCAAGATGTGTGTACGACAGCTTGATGCTAAAGACAGAGTCATG	1528
Dd	628	GlnProGlnSerAlaThrGlnGlyIleAlaTyrAspProLeuMetLeuLysHisGlnCysVal	647
Qy	1529	TGGCGGAAACACACAGCTGACCCCTGAACATGCTGGCGGATCCAGAGCATCTGCTCCGG	1588

D	b	648	CysGlyAsnSerThrTrpHisProGluHisAlaGlyIleGlnSerIleTrpArg	667
O	y	1589	CTCAGAGAGACAGCCCTGTTAGCAAGTGGAGACGGATCCGAGTGGCAACCCGCTA	1648
D	b	668	LeuGlnTrpTrpClyLeuLeuAsnLysCysGluArgIleGlnIleYArgLysAlaSerLeu	687
O	y	1649	GATGAGATCCAGACAGACGACTCGTAATACCAACACCCCTCTATAGGACAGTCCCTC	1708
D	b	688	GluGluIleGlnLeuValHisSerGlnHisHisSerLeuLeuTrpClyTrpHisProIleu	707
O	y	1709	AACCGGCAAGACTAGACAGCAAGAAAGTTGCTCGTCCATCAGCCAGAAAGTATGCT	1768
D	b	708	AspGlyGlnLysLeuAspProArgIleLeuLeuGlnLYAspAspSerGlnLysPheHeSer	727
O	y	1769	GTGTCGCTTGTGGGGGACCTGGGGTGGAGACAGACCCGTGGGAATGAGATGGACCTC	1828
D	b	728	SerLeuProCysGlyGlyLeuGlnLYAlaAspSerAspTrpHisIleTrpAsnIleuHisSer	747
O	y	1829	TCCAGTGTGTGCGCATGGCAGATGGGTGGCTGCTGTGAGCTGGACCTTCAAGGTGGCTCA	1888
D	b	748	SerGlyAlaAlaArgMetAlaValAlaGlyValIleGlnIleuAlaSerLYValAlaSer	767
O	y	1889	GGAAGAGCTCAAGATGGATTTGGCCATCATCCGGCCCCCAGACACCCCGGAGAAATCC	1948
D	b	768	GlyClyLeuLysAsnGlyPheAlaValAlaArgProProGlyHisHisAlaGlnIleuSer	787
O	y	1949	ACAGCCATGGATTTCTGCTCTTCAACTGTGAGCCATCAGCCGAAATCCTTACAGCAG	2008
D	b	788	ThrAlaMetGlyPheCysPhePheAsnSerValAlaIleIleThrAlaLysTrpLeuArgSp	807
O	y	2009	AAGTTGAACGTGGGCAAGCTCTCATCTGTGAGCTGGAGCAATTCACCATGGCAATGGCAC	2068
D	b	808	GlnLeuAsnIleSerLysIleLeuIleValAlaAspLeuAspValHisHisGlyAsnIleYTr	827
O	y	2069	CAGCAGGGGTTCTACATGACCCCTGATGCTCTTAATCTGTCGTGATGAGTATGACAAAC	2128
D	b	828	GlnGlnAlaPheTrpAlaAspProSerIleLeuTrpIleSerLeuHisArgTrpArgIleu	847
O	y	2129	GGGAATTTCTTCCAGGCTGTGGGGTCTCGAAGAGTTGTGGAGACCAAGCTGTGGGG	2188
D	b	848	GlyAsnPhePheProGlySerGlyAlaProAsnGluValAlaGlyHisClyLeuGlyGlyCly	867
O	y	2189	TACACATGACGTCGGATGGACAGACAGAGTCGTGACCCGCCCATTGGAGAGCTGGAGTAC	2248
D	b	868	TyrAsnIleAsnIleAlaTrpTrpGlyGlyLeuAspProProMetGlyAspValGluTrp	887
O	y	2249	CTTACAGCTTCAGAGACAGTGTGATGTGCCATTTGCCAGAGCTTGCATCTATGTGCTC	2308
D	b	888	LeuGlnAlaPheArgTrpIleValLysProValAlaLYGluGlnPheAspProMetVal	907
O	y	2309	CTAGTCTCCGCGGGTTGATGCTGTTGAAGACATCTGTCTCTCTGGGTGGCTACTCT	2368
D	b	908	LeuValSerAlaGlyPheAspAlaLeuGlnGlyHisTrpProProLeuGlyIleTrpLys	927
O	y	2369	GTACACCGCAGATGTTTGGSCACTTGACACGACACACTATGACCCGTGGACGAGGGCCGG	2428
D	b	928	ValThrAlaLysCysPheGlnLysIleuThrLYGlnLeuMetThrLeuAlaSerIleArg	947
O	y	2429	GTGCTGCTGGCTGGAGGAGACGCGCATGACTTACGCCCATCTGTGATGCTCTTAAGCT	2488
D	b	948	ValValLeuAlaLeuGlnGlyGlyHisAspLeuThrAlaIleCysAspAlaSerGluAla	967
O	y	2489	TGTATCTCGGCTGTGCGTCAGTGTAAAGTCGACCCCTTGATGAGGACAGTCTTGACGAA	2548
D	b	968	CysValAsnAlaLeuLeuGlyAsnGlyuLeuGluProLeuAlaGlnAspIleuHisGln	987
O	y	2549	AAGCCCAACAATCAACGCACTGGCCGACGCTAGAGAAAGTTCATGAGATCCAGAC	2602
D	b	988	SerProAsnMetAsnAlaValIleSerLeuGlnLysIleIleGlnIleGlnSer	1005
RESULT 10				
US-10-072-094-93				
: Sequence 93, Application US/10072094				

```

: Publication No. US20030025538A1
: GENERAL INFORMATION:
: APPLICANT: JACKSON, DONALD
: APPLICANT: LORENZI, MATTHEW
: APPLICANT: ATAK, RICARDO
: APPLICANT: GOTTFARDS, MARCO
: TITLE OF INVENTION: NOVEL HUMAN HISTONE PEPTIDES
: FILE REFERENCE: 3053-4145US1
: CURRENT APPLICATION NUMBER: US/10/072, 094
: CURRENT FILING DATE: 2002-06-14
: PRIOR APPLICATION NUMBER: 60/298, 296
: PRIOR FILING DATE: 2001-06-14
: NUMBER OF SEQ ID NOS: 127
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 93
: LENGTH: 780
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-10-072-094-93

Alignment Scores:
Pred. No.: 2,27e-101 Length: 780
Score: 2212.50 Matches: 472
Percent Similarity: 71.06% Conservative: 105
Best Local Similarity: 58.13% Mismatches: 184
Query Match: 41.39% Indels: 52
Gaps: 17

US-09-502-945-2 (1-2885) x US-10-072-094-93 (1-780)
QY 425 GCTGAGATAGGCTTTACTGCTCAGTCCCGACATCCCG-----ACTGAGATGCTCCT 478
Db 1 Alagluasnlu---ThrservalLeupProThrProhlsAlagluInMetValSer 19
QY 479 CAGCACGAGACCCCTCCCTGACAGACCTCCCGACAGTTCAGCTTACAGCTCTCT 538
Db 20 GlnGlaArgleuLeuIleHisGluAspSerMetAsnLeuSerLeuYrThrSerPro 39
QY 539 TCTGCTCCCAACATCTCCAGGCTGACAGGCGACGCGTCACTGTACCAACATCACACTC 598
Db 40 SerLeuProAsnIleThrLeuGluLeuProAla-----ValProSerGlnLeu 55
QY 599 ACTGCTCCCGGAGAGCTGTGACACAGACAGAGCGGACGAGCGCTCCAGTCCCTG 658
Db 56 AsnAlaSerHisSerLeuYsgluYsglnYsglsu-----ThrGlnThrLeu 72
QY 659 CGGACGAGTGGACGCTGACCGGCAAGTTC-----ATGACACATCTCTATT 706
Db 73 ArgGlnGluYalProleuProGluGlnYglYSerIleProAlaSerSerHis 92
QY 707 CTTGGCTGCTGCTGGCGCTGGACATGAGGCGGAGCGGACCGCCCGAGGGCATGCTCC 766
Db 93 Pro-----HisValThrLeuGluGluYsProProAsnSerSerHisGlnAla 108
QY 767 CTGCTCAGACATGCTGTTGCTGAGACAGCGCGGACAGACAGACCCCTCATTTGCT 823
Db 109 LeuLeuGlnHisIleuLeuYsglnGlnMetArgGlnIleuLeuValAlaGly 128
QY 824 ---GTGCCACTCCAGCGGACGCTCCCACTAGTACGAGGTGAACGTGTGGCCACAGCATG 880
Db 129 GlnYalProleuHisProGlnSerProleuAlaThrlYsGlnArgIleSerProGluIle 148
QY 881 CGGACGCTGAGCAAGCTCCCGGCGCATCGCCCTGAGCGGCACTAGTCTTACCGCTG 940
Db 149 ArgGluThrHisIysLeuProArgHisArgProleuAsnArgThrGlnSerAlaProleu 168
QY 941 CCCGACAGTCCCGAGCGCTGACAGCTGTGATCAACACAGACCGGAGTTCCTG 1000
Db 169 ProGlnSer-----ThrlLeuAlaGlnLeuValIleGlnIleGlnHisGlnIleu 186
QY 1001 GAGAACGACAGAG-----CAGCAGACAGCTACAGCTGGGCAAGATCTTACCAAGACAG 1054
Db 187 GluYsglnYsglnYrGlnGlnGlnIleHisMetAsnYsLeuSerLeuYsSerIle 206

```

```

QY 1055 GAGCTGCCAGCGAGCGCCACCCACCGCTGAGAGACAGAGGAGGAGTGCAGGACGAG 1114
Db 207 GlnGluLeuYsglnProGlySerHisLeuGlnGlnAlaGluIleuLeuGln----- 224
QY 1115 CAGGAGGCTTCTGGGAGGAGGAGCCCTGACCATCGCCCGGAGGAGGCTCCACAGAGAGT 1174
Db 225 -----GlyAspGlnAlaMetGlnGlnIleAspArgAlaProSerSer---Gly 238
QY 1175 GAGACACACAGAGAACTCGGAGGAGGAGGAGGAGGAGAGAGATGGAGGAGGAGGAG 1234
Db 239 AsnSerThrArgSerAspSerSerAlaCysValAspAspThrLeuIly-----GlnVal 256
QY 1235 GATTGCATCCAGGTTTAAAGACGAGGAGGAGGAGGAGGAGTGTCTGATGCGGAGGAGGAG 1294
Db 257 GlyAlaValIysValIysGlnGlnIleProValAspSer-----AspGlnAspAlaGlnIle 274
QY 1295 GAGAGACCTGTGCTGGATACAAAAAATCTGTCTCAAGTCCCAACCGCTGCAAGCT--- 1351
Db 275 GlnGluMetGlnSerGlnGlnGlnAlaIleAlaPheMetGlnIleProPheLeuGlnProThr 294
QY 1352 -----TTGCAGGTGTACCAAGCGCCCTTACAGCTGGCGCACTGTG----- 1390
Db 295 HisThrArgAlaLeuSerValArgGlnAlaProleuAlaValAlaGlyMetAspGlyLeu 314
QY 1391 ---CCGACACAGACCCCGGCGCTGACCAATCTCCCGCTGCTGCTGCGGAGCATGAGAG 1447
Db 315 GluYsHisArgLeuValSerArgThrHisSerSerProAlaIleSerValIleProHis 334
QY 1448 AACCCCGGACAGCAACCCGCTGCAAGACACTCTTCCACCAAGTGTGCTTACAGCAAGTTC 1507
Db 335 ProAlaMetAspArgProleuGlnIleProGlySerAlaThrlYglYleAlaYrAspProleu 354
QY 1508 ATGCTAAAGCAGCAATGATATGTCGGGAGACACACAGCTGACACCTTACAGCTGTGCGCGG 1567
Db 355 MetLeuYsHisGlnYsValIysGlnYsAspSerThrThrHisProGlnHisAlaGlyArg 374
QY 1568 ATCCAGGACATCTGTGCTGCCGCTGACGAGACAGGCGCTGTGCAAGTCCGAGCGGAGTC 1627
Db 375 IleGlnSerIleThrPheArgLeuGlnGlnIleThrlYglYleuYsAsnYsGlnIleGlyIle 394
QY 1628 CGAGTTCGCAAAAGCAAGCTGATGAGATCCAGACAGTGCATCTGATATCCACACCTG 1687
Db 395 GlnGluYArgYsAlaSerLeuGlnIleGlnLeuValHisSerGlnHisSerLeu 414
QY 1688 CTGTATGGACAGCTCCCTCAACCGGACAGAACTGACAGCAAGATGTCGCTGCCCTCC 1747
Db 415 LeuYrGluYrHisAsnProleuAspArgGlnYsLeuAspProArgIleLeuLeuGlyAsp 434
QY 1748 ATCAGCAGAAAGATGATGCTGTGCTGCTGCTGCGGAGCATCGGGGAGGACAGTGCACAC 1807
Db 435 AspSerGlnYsPhePheSerSerLeuProCysGluYglYleuIleYalAspSerAspThr 454
QY 1808 GTGTGATGATGATGACACTCTCTCAGTGTCTCCGATGAGTGCAGTGGGCTGTGCTGAG 1867
Db 455 IleThrAsnIleuHisSerSerGlnYalAlaArgMetAlaValGlyCysValIleGln 474
QY 1868 CTGCTTCAAGGTGTGCTGACAGAGACTGACAGATGATGATTTGCTCATCATCGGCCCCCA 1927
Db 475 LeuAlaSerYsValAlaSerGlnGlnLeuYsAsnGlyPheAlaValAlaArgProPro 494
QY 1928 GAGACACAGCGGAGGAAATCCACAGGCAGGATGCTGCTTCAACTGTGAGCATG 1987
Db 495 GlnHisHisAlaGluGlnSerThrAlaMetGlnYpHeYsPhePheAsnSerValAlaIle 514
QY 1988 ACCGCAAAACTCTTACAGCAAGAAATTGAAGCTGGGAGGATCTCTGATGAGTGGAG 2047
Db 515 ThrAlaIysTrpLeuArgAspGlnLeuAsnIleSerYsIleuIleValAspLeuAsp 534
QY 2048 ATTACACATGGCAATGGACCCGACAGAGCGTGTACAAATGACCCCTGTGCTGTACAG 2107
Db 535 ValHisHisGlnYsGlnYrGlnGlnIleAlaPheYrAlaAspProSerIleLeuYrIle 554

```

OY	2108	TCTGTGATCGCTTATTGCACAGGGAACTTTCTTCAGGCTCCTGGGACTCCGAAGGTT	2167
Db	555	SerLeuHisArgIyrAspRglUglnAsnPheProGlySerCylAlaProAsnGluVal	574
OY	2168	GCTGAGAGCAACGCCGTGGGCTACAATGTGAACSTGGCATGACAGAGCTGTGACC	2227
Db	575	GlyHicrGlyLeuGlnIgUglnIyTyraSnLIleAsnLIleAlaTrpHicGlyGlyLeuAspPro	594
OY	2228	CCCATGTGGAGACGCGTAGTACCTTTAACGCCCTTGAGAGACGTGGTGTGGCCATTTCCGCAC	2287
Db	595	PrometGlyAspValGluIyTrtleuGlnAlaPheArgThrLileValIysProAlaIalys	614
OY	2288	GAGTTCTCACCTGAATGTGGTCTAgtTCCCGGCGGTTGATGTCTTGAAGACATCTG	2347
Db	615	GluPheAspProAspMetValIleuValISerAlaGlyPheAspAlAleuGlnUgIyHisThr	634
OY	2348	TTCTGCTGGGTGGCTCTCTGTGTCACCGCCAGATGTTTTGGCCACTTATCCACGAGCATG	2407
Db	635	ProProLeuGlnGlyIyTrIyLysValIThrAlaLysCysPheGlyHisIleuThrIySgInIeu	654
OY	2408	ATGACCCCTGGCAGGGGCGCGGTGGTGTGCGCCCTGGAGAGAGCCATGACTTACCGCC	2467
Db	655	MetThrIeuAlaAspRglArGValValIleuAlaIleuGlnUgIyGlyHisAspLeuThrAla	674
OY	2468	ATCTGTGTAATGCTCTGAAGCTTGTGTCTGGCCTCTGCTAGTGTAAAGCTGACGCCCTTG	2527
Db	675	IleCysAspAlaSerGltValIacYSvalASnAlaIleuGlnIyAsnGlnIeudnUlProIeu	694
OY	2528	GATGAGCAGACTCTTGTCAGCAAAGCCCCAACATCAACGCGAGTGGCCACGCTAGAGAAATG	2587
Db	695	AlaGlnAspIleIeuHisGlnSerProAsnMetAsnAlaValILleSerIeudnIlyStIle	714
OY	2588	ATCGAGATCCAGAGCAAACTGTGAGCTGTGTGCAGAGTTCGCCGTGGTGTGGCCGG	2647
Db	715	IleGluIleGlnISerIyStrIyTrpIyLSerValArgMetValAlaValaProArgGlyLys	734
OY	2648	TCCCTCGGAGAGGGGCCCAAGCAGAGGTGAGACCAAGAAGCCGAAT--GTAGACGCCATGGCC	2706
Db	735	AlaIeuAlaIagIyAlaGlnIeu-----GlnGlnGlnIyThrGltThrValSerIalIeuAla	752
OY	2707	TTTGCTGTGGTGGGGGCGCAACAGAGGCCCAAGCTTGGCGGACCCCGGAACACAGCCCACG	2766
Db	769	ThrIaIyGlnUprometGlnGlnUpProAlaIeu 780	
 RESULT 11 US-10-072-094-107 Sequence 107, Application US/10072094 Publication NO. US20030025538A1  GENERAL INFORMATION: APPLICANT: JACKSON, DONALD APPLICANT: LORENZI, MATTHEW APPLICANT: ATTAR, RICARDO APPLICANT: GOTTARDIS, MARCO TITLE OF INVENTION: NOVEL HUMAN HISTONE PEPTIDES FILE REFERENCE: 3053-4145US1 CURRENT APPLICATION NUMBER: US/10/072_094 CURRENT FILING DATE: 2002-06-14 PRIOR APPLICATION NUMBER: 60/298,296 PRIOR FILING DATE: 2001-06-14 NUMBER OF SEQ ID NOS: 127 SOFTWARE: PatentIn Ver. 2.1. SEQ ID NO 107 LENGTH: 1141 TYPE: PRT ORGANISM: Artificial Sequence FEATURE: OTHER INFORMATION: Description of Artificial Sequence: Consensus OTHER INFORMATION: sequence			

1	NAME/KEY: MOD_RES		
2	LOCATION: (2)		
3	OTHER INFORMATION: Unknown	consensus	marker
4	FEATURE:		
5	NAME/KEY: MOD_RES		
6	LOCATION: (4)..(7)		
7	OTHER INFORMATION: Unknown	consensus	marker
8	FEATURE:		
9	NAME/KEY: MOD_RES		
10	LOCATION: (15)..(16)		
11	OTHER INFORMATION: Unknown	consensus	marker
12	FEATURE:		
13	NAME/KEY: MOD_RES		
14	LOCATION: (21)..(24)		
15	OTHER INFORMATION: Unknown	consensus	marker
16	FEATURE:		
17	NAME/KEY: MOD_RES		
18	LOCATION: (26)..(27)		
19	OTHER INFORMATION: Unknown	consensus	marker
20	FEATURE:		
21	NAME/KEY: MOD_RES		
22	LOCATION: (29)..(30)		
23	OTHER INFORMATION: Unknown	consensus	marker
24	FEATURE:		
25	NAME/KEY: MOD_RES		
26	LOCATION: (35)..(37)		
27	OTHER INFORMATION: Unknown	consensus	marker
28	FEATURE:		
29	NAME/KEY: MOD_RES		
30	LOCATION: (40)..(41)		
31	OTHER INFORMATION: Unknown	consensus	marker
32	FEATURE:		
33	NAME/KEY: MOD_RES		
34	LOCATION: (43)..(46)		
35	OTHER INFORMATION: Unknown	consensus	marker
36	FEATURE:		
37	NAME/KEY: MOD_RES		
38	LOCATION: (49)..(62)		
39	OTHER INFORMATION: Unknown	consensus	marker
40	FEATURE:		
41	NAME/KEY: MOD_RES		
42	LOCATION: (66)..(67)		
43	OTHER INFORMATION: Unknown	consensus	marker
44	FEATURE:		
45	NAME/KEY: MOD_RES		
46	LOCATION: (70)		
47	OTHER INFORMATION: Unknown	consensus	marker
48	FEATURE:		
49	NAME/KEY: MOD_RES		
50	LOCATION: (74)		
51	OTHER INFORMATION: Unknown	consensus	marker
52	FEATURE:		
53	NAME/KEY: MOD_RES		
54	LOCATION: (82)		
55	OTHER INFORMATION: Unknown	consensus	marker
56	FEATURE:		
57	NAME/KEY: MOD_RES		
58	LOCATION: (84)		
59	OTHER INFORMATION: Unknown	consensus	marker
60	FEATURE:		
61	NAME/KEY: MOD_RES		
62	LOCATION: (86)		
63	OTHER INFORMATION: Unknown	consensus	marker
64	FEATURE:		
65	NAME/KEY: MOD_RES		
66	LOCATION: (93)		
67	OTHER INFORMATION: Unknown	consensus	marker
68	FEATURE:		
69	NAME/KEY: MOD_RES		
70	LOCATION: (104)		
71	OTHER INFORMATION: Unknown	consensus	marker
72	FEATURE:		
73	NAME/KEY: MOD_RES		
74	LOCATION: (104)		
75	OTHER INFORMATION: Unknown	consensus	marker
76	FEATURE:		
77	NAME/KEY: MOD_RES		
78	LOCATION: (104)		
79	OTHER INFORMATION: Unknown	consensus	marker
80	FEATURE:		
81	NAME/KEY: MOD_RES		
82	LOCATION: (104)		
83	OTHER INFORMATION: Unknown	consensus	marker
84	FEATURE:		
85	NAME/KEY: MOD_RES		
86	LOCATION: (104)		
87	OTHER INFORMATION: Unknown	consensus	marker
88	FEATURE:		
89	NAME/KEY: MOD_RES		
90	LOCATION: (104)		
91	OTHER INFORMATION: Unknown	consensus	marker
92	FEATURE:		
93	NAME/KEY: MOD_RES		
94	LOCATION: (104)		
95	OTHER INFORMATION: Unknown	consensus	marker
96	FEATURE:		
97	NAME/KEY: MOD_RES		
98	LOCATION: (104)		
99	OTHER INFORMATION: Unknown	consensus	marker
100	FEATURE:		
101	NAME/KEY: MOD_RES		
102	LOCATION: (104)		
103	OTHER INFORMATION: Unknown	consensus	marker
104	FEATURE:		
105	NAME/KEY: MOD_RES		
106	LOCATION: (104)		
107	OTHER INFORMATION: Unknown	consensus	marker
108	FEATURE:		
109	NAME/KEY: MOD_RES		
110	LOCATION: (104)		
111	OTHER INFORMATION: Unknown	consensus	marker
112	FEATURE:		
113	NAME/KEY: MOD_RES		
114	LOCATION: (104)		
115	OTHER INFORMATION: Unknown	consensus	marker
116	FEATURE:		
117	NAME/KEY: MOD_RES		
118	LOCATION: (104)		
119	OTHER INFORMATION: Unknown	consensus	marker
120	FEATURE:		
121	NAME/KEY: MOD_RES		
122	LOCATION: (104)		
123	OTHER INFORMATION: Unknown	consensus	marker
124	FEATURE:		
125	NAME/KEY: MOD_RES		
126	LOCATION: (104)		
127	OTHER INFORMATION: Unknown	consensus	marker
128	FEATURE:		
129	NAME/KEY: MOD_RES		
130	LOCATION: (104)		
131	OTHER INFORMATION: Unknown	consensus	marker
132	FEATURE:		
133	NAME/KEY: MOD_RES		

```

? LOCATION: (110)..(111)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (114)..(115)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (119)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (126)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (128)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (134)..(138)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (143)..(144)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (147)..(148)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (158)..(160)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (163)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (165)..(166)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (169)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (173)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (176)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (185)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (193)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (195)..(202)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (205)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (208)

```

```

? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (210)..(212)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (215)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (218)..(221)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (223)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (232)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (238)..(240)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES
? LOCATION: (243)..(245)
? OTHER INFORMATION: Unknown consensus marker
? FEATURE:
? NAME/KEY: MOD_RES

```

```

Alignment Scores:
Pred. No.: 5.44e-100 Length: 1141
Score: 2184.50 Matches: 472
Percent Similarity: 55.51% Conservative: 57
Best Local Similarity: 49.53% Mismatches: 405
Query Match: 40.86% Indels: 20
DB: 9 Gaps: 6

```

```

US-09-502-945-2 (1-2885) x US-10-072-094-107 (1-1141)
QY 2 GAATTCCTCTTGTGAG-----TCAAAGAGCCACACGCGGCTCAACCATTC 55
DB 189 Glupheleu***Lys*****AsnGly***AsnHis*** 208
QY 56 CTCACACAGCAGCCCAAAATGCTG-----GAGCCACCATGCTCTTGGACGAGT 109
DB 209 Val*****Prolys***TPTyr*****His***SerLeuAspGlnSer 228
QY 110 TCCCTCCCGCAGAGCGGCCCTGGAGCCCTCCCTACAAAGCTCTTGCGCTGG 169
DB 229 SerProPro***SerGlyProGly*****Sertyr*****Leu***Gly 248
QY 170 CCCTAGCAGACGAGACGACTTCCCTCCGCAAAACAGCTCTGNAACCACTTGAA 229
DB 249 *****AspAlaLysAspAspPheProLeuArgLysThrLysSerGluProAsnLeuLys 268
QY 230 GTGCGTTCAAGGCTAAACGAAGGTGGCTGAGCGGAGAGAGTCCCTCGCTGCG 289
DB 269 ValArgSerArgLeuLysGlnLysValAlaGlnArgSerSerProLeuLeuArgArg 288
QY 290 AAGGATGGGACGCTGTATTAGCACCCTTTAAGAAGAGCTTTAGATCAGAGTGGCGG 349
DB 289 LysAspGly***ValValThr*****LysLysArg****GluValThrGlyAlaGly 308
QY 350 CCTGGGCGCTGTCGCTGTAAACGAGCGCGGCTCCGCGGCGGCTCCACAGC 409
DB 309 ProGly*****Ser*****SerSerProGlySerGlyProSerSerProAsnAsn 328
QY 410 ---TCCACAGCAGCATCGCTGAGAAATGCGTTACTGGCTCAGTCCCAACATCCCACT 466
DB 329 *****GluAsn*****Pro***** 348

```





```
Db 1069 ILILIEGLULIE***Ser*****Gly 1088
QY 2645 CGGTCCCTCGAGGGCCCAAGCAGGTGAGACCGAAGAACCGAAT-GTGAACGCCANG 2703
Db 1089 ***SerLeu***GluLa*****GluGlu***GluThrValSerAlaMet 1108
QY 2704 GCCTTGCTGTGTGGGGGCGCAACAGGCCAGCTGGCGACCCCGGAACACAGCCCC 2763
Db 1109 Ala***LeuSer***** 1128
QY 2764 AGGCGCGCAGAGAGCCCATGAGACGACCTGCCCC 2802
Db 1129 Arg*****GluProMetGlu***GluPro***Leu 1141

RESULT 12
US-10-072-094-90
; Sequence 90, Application US/10072094
; Publication No. US2003002538A1
; GENERAL INFORMATION:
; APPLICANT: JACKSON, DONALD
; APPLICANT: LORENZIL, MATTHEW
; APPLICANT: ATTAR, RICARDO
; APPLICANT: GOTTARDIS, MARCO
; TITLE OF INVENTION: NOVEL HUMAN HISTONE PEPTIDES
; FILE REFERENCE: 3053-4145US1
; CURRENT APPLICATION NUMBER: US/10/072, 094
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298, 296
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 90
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-094-90

Alignment Scores:
Pred. No.: 2,29e-90 Length: 879
Score: 1988.00 Matches: 431
Percent Similarity: 69.06% Conservative: 89
Best Local Similarity: 57.24% Mismatches: 161
Query Match: 37.19% Indels: 72
DB: 9 Gaps: 19

US-09-502-945-2 (1-2885) x US-10-072-094-90 (1-879)
QY 2 GAATTCCTCTGTGAGATCA-----AAGAGCCCGACACAGCGCGCTCAACATTC 55
Db 149 GILPhelLeuSerLysSerAlaThrLysAspThrProThrAsnGlyLysAsnHisSer 168
QY 56 CTCGCCACAGCACCACAAATGCTGG-----GAGGCCACCATGCTTTGGACAGAGT 109
Db 169 ValSerArgHisProLysLeuTrpTyrThrAlaAlaHisHisThrSerLeuAspGlnSer 188
QY 110 TCCCTCCCGACAGGGCCCGCCCTGGAGCGCCCTCCACAAAGCTGGCTTGGCGG 169
Db 189 SerProLeuSer-----GlyThrSerProSerTyrLysTyrThrLeuProGly 205
QY 170 CCTACGACAGTGTGAGACGACTTCCCTCCGCAAAACAGACCTCTAAACCAATTGAA 229
Db 206 AlaGlnAspAlaLysAspAspPheProLeuArgLysThrLAserGluProAsnLeuLys 225
QY 230 GTGCGTTCAAGGCTAAACAGAAAGTGGCTGACGCGAAGAGAGCTCCCTGCGTGC 289
Db 226 ValArgSerArgLeuLysGlnLysValAlaGluArgArgSerSerProLeuLeuArgArg 245
QY 290 AAGAGGGGAGCTGTATTACACTTGAAGAGAGAGAGCTTGAAGAGAGTGGCGGG 349
Db 246 LysAspGlyAsnValValIlnrSerPheLysLysArgMetPheGluValIlnr----- 262
QY 350 CCGTGGGCGTGTCCGTGTGTAAACAGCGACCCGCGCTCCGGCCCGACCTCTCCCAACAGC 409
```

```
Db 263 -----GluSerValSerSerSerProGlySerLysProSerProAsn--- 279
QY 410 TCCACAGACACCATCGCTGAGAAATGGCTATTAGGCTCACTCCCAACATCCCACTGAG 469
Db 280 -----AsnGlyProThrGlySerValIlnrGluAsnGlnThrSer 292
QY 470 ATGCTCCCT-----CAGACCGAGCGCCCTCCCTG 499
Db 293 ValLeuProProThrProHisAlaGluGlnMetValSerGlnGlnIlnrGlyLeuIlnrHis 312
QY 500 GACAGCTCCCAACCAAGTTTCAAGCTCTTACAGCTCTCTTGTGCGCAACATCTCCCTA 559
Db 313 GluAspSerMetAsnLeuLeuSerLeuTyrThrSerProSerLeuProAsnIlnrThrLeu 332
QY 560 GGGCTGCAGGCGCAGCGTCACTGTACCAACTACACCTCACTGCTCCCGCAAGTGTGC 619
Db 333 GlyLeuProAla-----ValProSerGlnLeuAsnAlaSerAsnSerLeuLys 348
QY 620 ACACAGCAGAGAGCCCGAGAGGAGCGCCCTCCACTCCCTGGCGAGGGTGGCAGCTGACC 679
Db 349 GluLysGlnLysCysGlu-----ThrGlnThrLeuArgGlnGlyValProLeuPro 365
QY 680 GGCAGAGTTC-----ATGAGCACATCTCTATTCTGCTGCTGCTGGCGGTG 727
Db 366 GlyGlnTyrGlyLysThrIleProAlaSerSerHisPro-----HisVal 381
QY 728 GCACGTGAGGCGCAGCGGAGCCCGACGGCATGCTCCCTGCTGCAGCATGTGCTGTG 787
Db 382 ThrLeuGluGlyLysProProAsnSerSerHisGlnAlaLeuGlnHisLeuLeu 401
QY 788 CTGAGCAGGCGCGGAGCAGCAGCACCCTCATGCT-----GTGGCAGTCAAGGGGAC 841
Db 402 LysGlnGlnMetArgGlnGlnLysLeuLeuValAlaGlyLysValProLeuHisProGln 421
QY 842 TCCCACTACTGACGGGTGACGCTGTGGCCACCGACGACGAGTGAAGCAGCTCCG 901
Db 422 SerProLeuAlaThrLysGlnArgLysIleSerProGlyLysArgGlyThrHisLysLeuPro 441
QY 902 CGGCATCGCGCCCTGAGCGCCGACCTCACTCACTGCTGCGCGAGAGTCCCAAGGCGCTG 961
Db 442 ArgHisArgProLeuAsnArgHisGlnSerAlaProLeuProIlnr-----ThrLeu 459
QY 962 CAGCAGCTGTATGCAACACAGCAGCAGCAGCTGCTGAGAGAGAG-----CAG 1015
Db 460 AlaGlnLeuValIlnrGlnGlnGlnHisGlnIlnrPheLeuGlnLysGlnLysGlnTyrGln 479
QY 1016 CAGCAGCTTACAGCTGGGCAAGATCTTCACCAACAGAGGGAGCTGCCAGGACCCAC 1075
Db 480 GlnGlnIlnrHisMetAsnLysLeuLeuSerLysSerIlnrGlnLeuLysGlnProGly 499
QY 1076 ACCCAGCTTGAAGAGACAGAGAGAGAGCTACGAGCAGCAGAGAGGCTTGTGGGGAG 1135
Db 500 SerHisLeuGlnGlnValAlaGlnGlnGlnLeuGln-----GlyAsp 512
QY 1136 GAGAGCCCTGACATGCCCCGGAGGGGCTCCACAGAGAGTGAAGACACAGAGAGACTG 1195
Db 513 GlnAlaMetGlnLysAspArgAlaProSerSer---GlyAsnSerThrArgSerAspSer 531
QY 1196 GAGAGAGAGAGCAGAGAAAGATGGGAGAGAGAGAGAGATTCATGCAAGTTAAAGAG 1255
Db 532 SerAlaCysValAspAspThrLeuGly-----GlnValGlyLysValLysValLysGln 549
QY 1256 GAGGAGGCGCAGAGTGTGCTGAGAGAGGGCCGCACTTGAAGAGAGCTGTGCTGATAC 1315
Db 550 GluProValAspSer-----AspGlnAspAlaGlnIlnrGlnIlnrMetGlnSerGlyGln 567
QY 1316 AAAAAAGTGTTCAGATGCCCAACCGCTCAAGCT-----TTGACAGGTG 1360
Db 568 GlnAlaAlaPheMetGlnGlnProPheLeuGlnProThrHisThrArgAlaLeuSerVal 587
QY 1361 TACCAAGCGCCCTCAAGCCTGAGCAGCTG-----CCGCAACCAAGCCCTGGCG 1408
```





Percent Similarity:	53.47%	Conservative:	101
Best Local Similarity:	42.34%	Mismatches:	195
Query Match:	31.27%	Indels:	227
DB:	10	Gaps:	22

QY	11	TTGTGAAGTCAAAAGGACCCACACAGAGGGGCGCTCAACATTCGTCGCCACGACACCC	70
Db	101	LeuGluAlaGlnValHisProAsnSerProGlyIleProIlyrAlaGlnIleuGluProIle	120
QY	71	AAATGCTGGGAGACCCACCATGCTTTCTTGACCAAGATTCCTCCGCCAGAGCGGCC	130
Db	121	GluThrGlnGlyAlaThrArgSerMetLeuSerSerPheLeuProProValProSerIle	140
QY	131	CCTGGAGCGCTCCCTCTCAACACTGGCTTTGGCTGGGCGCTACGACAGTCGAGAC	150
Db	141	ProSerAspProProGluHis-----	147
QY	191	TTCCCCCTCCCAAAACAGCCTCTGACACCCCAACTTGAAGTGCCTCAAGGCTAAACAG	250
Db	148	PheProLeuAlaGlySerThrValSerGluProAsnIleuLeuAlaGlyIlyProIys---	166
QY	251	AAGTGCGCTGACCGGACAGACAGTCCCTCTCGCTGGCAAGATGGGACTGTATTAGC	310
Db	167	LysSerLeuGlnAlaArgIlyAsnProLeuLeuAlaGlySerAlaPro-----Pro	184
QY	311	ACCTTTAAGAGAGGCGTGTGACATCACAGTGCAGCGGCGCTGGGCGCTGCCTGGTGT	370
Db	185	SerLeuAlaGlnAlaGlyProAlaGlnIleuThrLeuAlaGlyAspSerSerProSerSer	202
QY	371	AACAGCGCAGCGGCTCCGCGCCAGCTCTCCCAACAGTCCACAGACACATCGCTAG	430
Db	203	---SerThrProAlaSerGlyCysSerSerProAsnAspSer-----Glu	216
QY	431	AATGGCTTACTGGCTAGTGCCTCCCAACATCCCACTAGATGGTCCCTCAGACCGAGCC	490
Db	217	HisGly-----	218
QY	491	CTCCCTGTGACAGCTCCGCCAACAGTTCAGCCTTACAGGCTCTCTTCTGCCAAC	550
Db	219	-----ProAsn	220
QY	551	ATTCCTCCAGGGCGTGCAGGCGCACGGTACATCTACCAACTCACACCTGACTGCCCTCCG	610
Db	221	ProIleuGly-----	224
QY	611	AAGGTGTGCACACAGAGAGGCCGAGAGCGAGCGCTCCAGTCCCTGCGGCAAGGTGC	670
Db	225	-----AspSerAlaArgAlaGlnHisProIlyrIleuGlyProAlaGly	238
QY	671	ACGCTGACCGGCAAGTTCATGACACATCTCTATTCTT-----	709
Db	239	ProIleuGlySerProHisThrProLeuPheLeuProHisIlyLeuGluProGluAla	258
QY	710	---GAGTGCCTGCTGGGCGTGGGCACTGGAGGGCGACGGAGGCCCGGAGGATGGCTCC	766
Db	259	GlyIlyGlySerLeu-----ProSer	264
QY	767	CTGTGTCAGCATGTCTGTTCCTGGAGCAGGCCGCGGACAGACACACCTATTCGTGTG	826
Db	265	ArgLeuGlnProIleLeuIleLeuAspProSerGlySerHisAlaProLeuAlaThrVal	284
QY	827	CCACTCCAGGGGACGTCCSCCACTAGTACAGGGTAACTGTGTGGSCACACAGATCGGAGC	886
Db	285	ProGlyLeuGlyProLeuProPheHisPheAlaGlnSerIleMetThrGlnAlaGlyLeu	304
QY	887	GTAAGCAAGTCCCGGCGATCGGCGCTGAGCGCACTAGTCTCAACGCTGCGGCGAG	946
Db	305	SerIlySer---GlyLeuHisThrProLeuSerAlaGlnArgSerGlyThrProProPro	323
QY	947	AGTCCCAAGGCC-----CTGCAGCAGTGGTGCATG	976

Db	324	SerialThrAlaProProProProGlyProMetGlnProArgLeuIleu-----	341
Qy	977	CAACAACAGCACACAGACTTCTGGAGACAGAAACAGACAGCAGCTACAGCTGGGCAAG	1036
Db	342	---LysThrHisValGlnValIleLeuArgSer-----	351
Qy	1037	ATCCACACACAGACAGGGGAGACTCCAGGCAAGCCACACACACCTTGAGAGACAGAG	1096
Db	352	-----AlaLysProSerGluLysProArg-----	359
Qy	1097	GAGGAGCTGAGGAGACAGACAGAGAGCTTCTGGGGAGAGAGCCTGACATGCCCGG	1156
Db	359	-----	359
Qy	1157	GAGGCTCCACAGAGAGTGAAGACACACAGAAACCTGGAGAGAGAGACAGAGAGAG	1216
Db	360	-----LeuArgGlnIleProSerAlaGluProLeuGluThr	371
Qy	1217	GATGGGAGAGAGAGAGAGAGATGCTACAGATTAAAGACAGAGAGGCGGACAGTGT	1273
Db	372	AspGly-----GlyIleProGluGln	378
Qy	1274	---GCTGGAGAGGGCCGAGCTTGGAGAGCCTGGAGCTGGATACAAAAAGCTGTCTCA	1330
Db	379	ValValAspAspGlyLeuGlnHisArgGlnLeuGlnHisGlyGln-----Pro	394
Qy	1331	GATGCCACACCGCTGCACAACTTTGGCAGGTATACCAAGCGCC-----	1372
Db	395	GluAlaArgGlyProAlaProLeuGln-----GlnHisProGlnValIleLeuThrGlu	412
Qy	1373	-----CTCACG	1378
Db	413	GlnGlnArgLeuAlaGlyArgLeuProArgGlySerThrGlyAspCysValIleLeuPro	432
Qy	1379	CTGGCCACTGTGGCCCAAGAGCCCTGGGGCGTACCAACATCTCCCTCTGCGCCCTGAG	1438
Db	433	LeuAlaGlnGlyGlnHisAspProLeuSerArgAlaGlnSerProAlaIleProAla	452
Qy	1439	GGCATGAAGAACCC-----CCAGACCAACCC	1465
Db	453	SerIleSerAlaProGluIleProAlaSerAlaArgValLeuSerSerGluThrPro	472
Qy	1466	GTCAGGACACCTC---TTCACACACAGTGTGGTCTACGACACGTTACTGTAAGACACAG	1522
Db	473	AlaArgThrLeuProPheLeuThrGlyLeuIleLeuAspSerValMetLeuLysHisGln	492
Qy	1523	TGCATGTCCGGGACACACACAGCTGACCTTGACATGTCTGGCGCGATGCCAGAGCATCTGG	1582
Db	493	CysSerCysGlyAspAsnSerArgHisProGlnHisAlaGlyArgGlnIleSerIleThr	512
Qy	1583	TCCCGGCTGCAGAGACAGGCTCTTACCAATGGCGAGCGAGATCCGAGTGGCAAGCC	1642
Db	513	SerArgLeuGlnGlnIleArgGlyLeuArgSerGlnCysGlnIleAlaArgGlyArgLysAla	532
Qy	1643	ACGCTAGATGAGATCCAGACAGTGCACCTGTAATACCACACCTGCTATATGGACCACT	1702
Db	533	SerIleGluGlnLeuGlnSerValHisSerGluArgHisValLeuLeuLeuGlyThrAsn	552
Qy	1703	CCCCCAACCGGCGAAGGTAAAGACAGAAACAAATCTCTGGTCCGCAATCCAGCAAGAAAG	1766
Db	553	ProLeuSerArgLeuLysLeuAspAsnGlyLysLeuAlaGlyIleLeuAlaGlnArgMet	572
Qy	1763	TATGCTGTGCTGCTTGTGGGGCATCGGGGTGGACAGTACACCGTGTGCAATACAGATG	1822
Db	573	PheGluMetLeuProCysGlyGlyValGlyValAspThrHisPheThrIleLeuAsnIleu	592
Qy	1823	CACCTCTCCAGCTGTGGCATGAGAGTGGGCTGGCTGTGAGAGCTGGCTTCAAGGTG	1882
Db	593	HisSerSerAsnAlaIleArgTrpAlaIleArgLysSerValThrAspLeuAlaPheLysVal	612
Qy	1883	GCTGCAGAGAGCTCAAGATGATTTGCCATATCTCGGCCCCCAGAGACACAGCCGAG	1942
Db	613	AlaSerArgGluLeuLysAsnGlyPheAlaValValArgProProGlyHisHisLasp	632

```

QY 1943 GAATCCAGACCCATGGATTCCTGCTTCACCTGTACCATCCGCAAAACTCCTA 2002
DB 633 HisSerThrAlaMetGlyPheCysPhePheAsnSerValAlaIleAlaCysArgGlnLeu 652
QY 2003 CACGACAG------TTGAACGTGGGCAAGGTCCTCATCTGAGTCGGACATTCAC 2053
DB 653 GlnGlnGlnSerLysAlaSerLysAlaSerLysIleLeuIleValAspTrpAspValHis 672
QY 2054 CATGGCAATGGACCCACGACGGGCTTCACAAATGACCCCTGCTGTACATCTCTG 2113
DB 673 HisGlyAsnGlyThrGlnGlnThrPheTyrglnAspProSerValLeuTyrlleSerLeu 692
QY 2114 CATGGCTATGACAGGAGAACTTCCTTCAGGCTCGGAGCTCCTGTAAGAGGTGGGGA 2173
DB 693 HisArgHisAspAspGlyAsnPhePheProGlySerGlyAlaValAspGlnValGlyAla 712
QY 2174 GGACCAAGCGCTGGGATCAATGTGAACGTGGCATGACAGAGAGGTGGAACCCCATTC 2233
DB 713 GlySerGlyGlnGlyPheAsnValAsnValAlaIleTrpAlaGlyLeuAspProMet 732
QY 2234 GGAGAGGTGGAGTACCTTACAGCCTTCAGCAGCAGTGTGAGTGGCCATGGCCACGATTC 2293
DB 733 GlyAspProGlnTrpLeuAlaAlaPheArgIleValAlaMetProIleAlaArgGlnPhe 752
QY 2294 TCACCTGATGTGGTCTAGTCTCCGCCGGGTTGATGCTGTGAAGACATCTGTCTCT 2353
DB 753 SerProAspLeuValLeuValAsnIleArgIlePheAspAlaIleGlnGlyHisProAlaPro 772
QY 2354 CTGGGTGGCTACTCTGTACCGCCAGATGTTGGCCACTTGACCAAGGACGTGATGACC 2413
DB 773 LeuGlyGlyTyrlleHisValSerAlaIleCysPheGlyTyrlleMetThrGlnGlnLeuMetAsn 792
QY 2414 CTGGCAGGGGGCGGGGTGGTGGTGGCCCTGGAGGAGGCCATGACTTGAACCGCATCTGT 2473
DB 793 LeuAlaGlyGlyAlaValAlaValLeuAlaLeuGlnGlyGlyHisAspLeuTrpAlaIleCys 812
QY 2474 GATGCTCTGAACCTGTGCTGCTGGCTGTGCTGCTGCTGAAGTGAAGCCCTTGATGAG 2533
DB 813 AspAlaSerGlnAlaCysValAlaAlaIleLeuLeuGlyAsnArgValAspProLeuSerGln 832
QY 2534 GCAGTCTTGACGCAAAAGCCC 2554
DB 833 GlnGlyTrpLysGlnLysPro 839

RESULT 15
US-09-817-538-13
; Sequence 13, Application US/09817538
; Patent No. US20020137162A1
; GENERAL INFORMATION:
; APPLICANT: LI, Zuomel
; APPLICANT: Bonfils, Claire
; APPLICANT: Besterman, Jeffrey
; TITLE OF INVENTION: Antisense Oligonucleotide Inhibition of Specific Histone
; FILE REFERENCE: 106101.144
; CURRENT APPLICATION NUMBER: US/09/817,538
; CURRENT FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: US 60/192,157
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Human
US-09-817-538-13

Alignment Scores:
Pred. No.: 7.43e-75 Length: 855
Score: 1671.50 Matches: 384
Percent Similarity: 53.47% Conservative: 101
Best Local Similarity: 42.34% Mismatches: 195

```

```

Query Match: 31.27% Indels: 227
DB: 10 Gaps: 22
US-09-502-945-2 (1-2885) x US-09-817-538-13 (1-855)

QY 11 TTGTGCAAGTCAAAAGAGACCCACACAGAGGCGGCTCAACCATTCCTCCACAGACACCC 70
DB 101 LeuGlnArgThrValHisProAsnSerProGlyIleProIleArgThrLeuGlnProIle 120
QY 71 AAATGCTGGGGAGACCCACCAATGCTTCTTTGGACACAGATTCCCTCCACAGAGGCGCCC 130
DB 121 GlnThrGlnGlyAlaIleThrArgSerMetLeuSerSerPheLeuProProValProSerIle 140
QY 131 CGGGAGCGCTCCCTCCCTCAAAAGTCCCTTTCGCGGCGCTACAGACACTCGAGACAG 190
DB 141 ProSerAspProProGlnHis----- 147
QY 191 TTCCTCCCTCCGCAAAAGACCTGTGAACCCCACTTGAAGTGGCTTCMAAGCTAAACAG 250
DB 148 PheProLeuArgLysThrValSerGlnProAsnLeuLysLeuArgTyrlleProLys--- 166
QY 251 AAGTGGCTAGGGGAGAACACAGTCCCTCTGCTGCGCAAGATGGAGTGTATTAGC 310
DB 167 LysSerLeuGlnArgArgLysAsnProLeuLeuArgLysGlnSerAlaPro-----Pro 184
QY 311 ACCTTTAAAGAGAGAGCTGTGAGATCACAGGTGCGGCGCTGCGGCTGCTGCTGT 370
DB 185 SerLeuArgArgArgProAlaGlnThrLeuGlyAspSerSerProSerSer----- 202
QY 371 AACAGCGACCCGGGCTCCGCGCCAGCTCTCCCAACAGTCCACAGACACCATCGCTGAG 430
DB 203 ---SerThrProAlaSerGlyCysSerSerProAsnAspSer-----Gln 216
QY 431 AATGGCTTACTGAGTCAATGCCCAACATCCCACTAGATGCTCCTCAGACACGAGCC 490
DB 217 HisGly----- 218
QY 491 CTCCTCTGACAGATCCGCCCAACAGTTACAGCTGTACAGTCTCTCTGCCCCAAC 550
DB 219 -----ProAsn 220
QY 551 ATCTCCCTAGGCTCGAGGCGCACGGTCACTGTACACCAACTCAGACACTGAGCTCCCGC 610
DB 221 ProIleLeuGly----- 224
QY 611 AAGCTGTGACACAGACGAGGCGGAGAGGCGCCCTCCAGTCCCTCGCGGAGGGTGGC 670
DB 225 -----AspSerAspArgArgThrHisProThrLeuGlyProArgGly 238
QY 671 ACGCTGACGGCAAGTTCATGAGCAGACATCCTATTCCT----- 709
DB 239 ProIleLeuGlySerProHisThrProLeuPheLeuProHisGlyLeuGlnProGlnAla 258
QY 710 ---GGTGGCTGTGGGCGGTGGCACTGGAGGGGAGGAGGCGCCACAGGGCATGCTCC 766
DB 259 GlyGlyCysLeu-----ProSer 264
QY 767 CTGCTGACAGATGCTGTTGCTGTGAGACAGGCGGCGGAGACAGACCCATTCGCTGTG 826
DB 265 ArgLeuGlnProIleLeuLeuAspProSerGlySerHisAlaProLeuLeuThrVal 284
QY 827 CCACTCCAGCGGCAAGTCCCACTAGTGAAGGAGGAGCGTGGCCACAGACATGCGGAGAG 886
DB 285 ProGlyLeuGlyProLeuProPheHisPheAlaGlnSerIleMetThrTrpGlnValGln 304
QY 887 GTAAGCAAGCTCCGCGGCACTGCGGCGGAGGCGGAGCTGCTACAGCGGCTGCGGAG 946
DB 305 SerGlySer---GlyLeuHisTrpProLeuSerArgThrArgSerGlnProLeuProPro 323
QY 947 AGTCCCGAGGCG-----CTGACAGACAGCTGTGATG 976
DB 324 SerAlaThrAlaProProProProGlyProMetGlnProArgLeuGlnGlnLeu----- 341

```





GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 21, 2003, 12:34:32 ; Search time 22.35 Seconds

(without alignments)  
7595.986 Million cell updates/sec

Title: US-09-502-945-2

Perfect score: 5346  
Sequence: 1 ggaattcccttgcagagt.....aaacaaagaattaaaattt 2885

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xlp  
-O=/cgn2\_1/USFTO.spool/US0502945/runcat\_14032003\_101059\_19145/app\_query.fasta\_1.10979  
-DB=Issued\_Patents\_AA -OFMT=fasta -SUFFIX=ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosun62 -TRANS=human40.cd1  
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=US0502945 @CGN\_1.1.80 @unat\_14032003\_101059\_19145 -NCP=6 -ICPI=3  
-NO\_XLXLY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA: \*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/1/iaa/6C\_COMB.pep: \*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	374.5	7.0	335	4	US-09-446-504-80
2	374.5	7.0	335	4	US-09-712-266-80
3	302	5.6	482	1	US-08-528-255A-1
4	302	5.6	482	1	US-08-717-365-1
5	278	5.2	517	4	US-09-282-305-8
6	277.5	5.2	458	4	US-09-282-305-2
7	275.5	5.1	2088	4	US-09-548-372D-13
8	275.5	5.1	2088	4	US-09-548-367D-13
9	271.5	5.1	439	4	US-09-282-305-6
10	270.5	5.1	351	4	US-09-282-305-1
11	265	4.9	1400	4	US-08-630-915A-37
12	262.5	4.9	1917	4	US-09-627-650B-5

c 13	262.5	4.9	1917	4	US-09-436-063C-5	Sequence 5, Appl1
c 14	260	4.9	432	4	US-09-282-305-10	Sequence 10, Appl1
c 15	254	4.7	2601	4	US-09-627-650B-9	Sequence 9, Appl1
c 16	254	4.7	2601	4	US-08-436-063C-9	Sequence 9, Appl1
c 17	253	4.7	2211	4	US-09-738-884-1	Sequence 1, Appl1
c 18	248.5	4.6	2544	4	US-09-627-650B-3	Sequence 3, Appl1
c 19	248.5	4.6	2544	4	US-09-436-063C-3	Sequence 3, Appl1
c 20	241	4.5	1652	4	US-09-627-650B-1	Sequence 1, Appl1
c 21	241	4.5	1652	4	US-09-436-063C-1	Sequence 1, Appl1
c 22	239	4.4	2508	4	US-09-627-650B-7	Sequence 7, Appl1
c 23	239	4.4	2508	4	US-09-436-063C-7	Sequence 7, Appl1
c 24	235	4.4	1958	1	US-07-945-283-2	Sequence 2, Appl1
c 25	234	4.3	1958	1	US-07-945-283-2	Sequence 2, Appl1
c 26	231.5	4.3	960	4	US-09-219-849-6	Sequence 6, Appl1
c 27	229.5	4.3	714	4	US-08-556-978B-61	Sequence 61, Appl1
c 28	229.5	4.3	714	4	US-09-247-806-10	Sequence 10, Appl1
c 29	227.5	4.3	1065	1	US-08-642-255-72	Sequence 72, Appl1
c 30	224.5	4.2	882	4	US-09-413-814-78	Sequence 78, Appl1
c 31	224	4.2	720	4	US-09-219-849-4	Sequence 4, Appl1
c 32	224	4.2	777	1	US-08-642-255-53	Sequence 53, Appl1
c 33	222.5	4.1	1417	4	US-08-900-230-3	Sequence 3, Appl1
c 34	222	4.2	960	4	US-09-219-849-5	Sequence 5, Appl1
c 35	221.5	4.1	1128	4	US-09-627-650B-11	Sequence 11, Appl1
c 36	221.5	4.1	1128	4	US-09-436-063C-11	Sequence 11, Appl1
c 37	219.5	4.1	1057	3	US-08-931-820-1	Sequence 1, Appl1
c 38	219.5	4.1	1345	2	US-08-977-767-3	Sequence 3, Appl1
c 39	217.5	4.1	1461	4	US-09-585-887-9	Sequence 9, Appl1
c 40	217.5	4.1	1461	4	US-08-289-578-9	Sequence 9, Appl1
c 41	215.5	4.0	762	1	US-08-642-255-120	Sequence 120, App
c 42	215.5	4.0	762	1	US-08-397-633A-31	Sequence 31, Appl1
c 43	211	3.9	801	1	US-07-906-349A-6	Sequence 6, Appl1
c 44	207	3.9	2088	4	US-09-548-372D-13	Sequence 13, Appl1
c 45	207	3.9	2088	4	US-09-548-367D-13	Sequence 13, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-446-504-80  
Sequence 80, Application US/09446504  
Patent No. 6218150  
GENERAL INFORMATION:  
APPLICANT: UEMORI, Takashi  
APPLICANT: SATO, Yoshimi  
APPLICANT: FUJITA, Tomoko  
APPLICANT: MIYAKE, Kazuo  
APPLICANT: MURAI, Hiroyuki  
APPLICANT: ASADA, Kiyozo  
APPLICANT: KATO, Ikunoshin  
TITLE OR INVENTION: DNA POLYMERASE-RELATED FACTORS  
FILE REFERENCE: 1422-408PCT  
CURRENT APPLICATION NUMBER: US/09/446, 504  
CURRENT FILING DATE: 1999-12-23  
PRIOR APPLICATION NUMBER: PCT/JP98/02845  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: JP 9-187496  
PRIOR FILING DATE: 1997-06-26  
PRIOR APPLICATION NUMBER: JP 9-320692  
PRIOR FILING DATE: 1997-11-27  
NUMBER OF SEQ ID NOS: 92  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 80  
LENGTH: 335  
TYPE: PRT  
ORGANISM: Pyrococcus furiosus  
US-09-446-504-80  
Alignment Scores:  
Pred. No.: 1.27e-17  
Score: 374.50  
Percent Similarity: 43.98%  
Best Local Similarity: 28.27%  
Query Match: 7.01%  
Length: 335  
Matches: 108  
Conservative: 60  
Mismatches: 121  
Indels: 93

```

DB: 4 Gaps: 16
US-09-502-945-2 (1-2885) x US-09-446-504-80 (1-335)
QY 1547 CACCTGAGATGCTGGCCGGATCCAGAGATGCTGGCCGGCGGCGAGAGAGAGCGCTG 1606
DB 20 HisProGluAsnProLysArgLeuGluIleValLeuSerLysValArgGluLeuGlyLeu 39
QY 1607 CTAGCAAGTGCAGCGAGATCCAGATCCGACCAAGCCAGCTAGATGAGATCCAGAGCTG 1666
DB 40 GluGluArgIleGluGluProAsnProValGluGluThrPhe-----ValGluLysIle 57
QY 1667 CACTGTGAATACCAACCCCTCTCTATGGACAGTCCCTCAACCGGCAAGAGTAGAC 1726
DB 58 HisAspArgAspTyr-----IleAsnPheValLysGluAla 69
QY 1727 AGCAAGAAGTGTGGCTGCCATCAGCCAGAAAGATGATATGCTGTGGCTGTGGGGG 1786
DB 70 ValGluLysGlyIleThrArgLeuAspProAspThrTyrVal----- 83
QY 1787 ATCGGGGTGGACAGTACACCGGTGGAGATGAGATGCATCTCCAGTGTGGCGATG 1846
DB 84 -----SerProGlyThrTrp-----SerAlaIleLeuLeu 93
QY 1847 GCAGTGGGCTGCTGCTGGAGCTGGCTTCAAGTGGCTCCAGAGAGCTCAAGATGGA 1906
DB 94 AlaLeuGlyAla-----AlaArgSerAlaIleLeuSerAlaLeuHisTyrGly 109
QY 1907 -----TTTGCATCAATCGGCGCCCGAGACACCGCGCGAGAA----- 1945
DB 110 GlyLeuHisMetAlaLeuValArgProProGlyHisHisAlaGlyArgGlyAlaGlyAla 129
QY 1946 -----TCCACAGCCATGGAGATTCTGCTTCAACTCTGTACCATCACCGGAAAACTC 1999
DB 130 MetGlyAlaProThrLeuGlyPheCysIlePheAsnAsnAlaIleSerAlaValIleThr 149
QY 2000 CTACAGCAAGAGTGAACGTGGGCAAGTCTCATGCTGAGTGGAGCATTCACCATGCG 2059
DB 150 LeuLysGluLeu---GlyValGlyLysValValIleAspPheAsnAlaHisGly 168
QY 2060 AATGGACCCAGAGAGGCTCTACAAATGACCCCTGTGCTACATCTGTGATCGC 2119
DB 169 AsnGlyThrClnIleIlePheTrpAsnAspProAspValIleHisIleAspLeuHisGly 188
QY 2120 TATGACAAAGGAACTTTTCCAGGCTTGCGGCTCCTGAAGAGTGGTGAGAGACA 2179
DB 189 ArgAsp-----IleTyrProGlySerGlyAspValSerGlyValGlyLysSerAsn 205
QY 2180 GCGGTGGGTACATGTGAACGTGGCATGAGACAGAGGTGTGACCCCATTT----- 2233
DB 206 AlaTyrGlySerLysIleAsnLeu-----PrometProHis 217
QY 2234 -----GGAACGTGGAGTACCTTACAGCCTTCAGACATGTGGATGCCATTTGCCAC 2287
DB 218 TyrSerGlyAspLysAspTyrIleTyrValTrpAspGluIleValLeuProIleValGlu 237
QY 2288 GAGTTCTACCATGTGTGCTAGTCTCCGCGGGTTGATCCTGTGTAAGACAT--- 2344
DB 238 GluValLysProLysValIleValIleSerAlaGlyPheAspGlyPheLysGlyAspGly 257
QY 2345 CTGTCTCCTCTGGGTGCTACTCTGCACCGCCAGATGTTTGGCCACTTGACACAGGAG 2404
DB 258 LeuThrThrLeu-----ArgLeuThrGluSer 266
QY 2405 CTGATACCCCTGGCAGGAGGCGCG-----GTGGTGTGGCGCTGGAGGGA 2449
DB 267 PheTyrSerTyrAlaGlyAlaThrLeuAsnLysTyrProLeuAlaPheIleLeuGluGly 286
QY 2450 GGGCATGACTTG-----ACCGCATCTGTGATGCGCTGAGAGCT 2488
DB 287 GlyTyrGlyValGlyLeuAspLysGlyPheProAlaPheIleMetGlyTyrGluGluGly 306
QY 2489 TGTGTCGGCTCTGCTCAGTGAAGCTGCACGCTTGATGAGGCACTTTCGAGCAA 2548

```

```

DB 307 -----LysAlaLysAlaArgGlu 312
QY 2549 AAGCCCAACATCAACGAGGCGCCAGCTAGACAAGTCAATGACATCCAGACAAACAC 2608
DB 313 GluProArgTyrGluThrLeuLysLeuAlaGluGluValLysAspIleLeuSerProTrp 332
QY 2609 TGGAGC 2614
DB 333 TrpSer 334

RESULT 2
US-09-712-266-80
; Sequence 80, Application US/09712266
; Patent No. 633158
; GENERAL INFORMATION:
; APPLICANT: UEMORI, Takashi
; APPLICANT: SATO, Yoshimi
; APPLICANT: FUJITA, Tomoko
; APPLICANT: MIYAKE, Kazuo
; APPLICANT: MUKAI, Hiroyuki
; APPLICANT: ASADA, Kiyoko
; APPLICANT: KATO, Kunoshin
; TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS
; FILE REFERENCE: 1422-408PCT
; CURRENT APPLICATION NUMBER: US/09/712,266
; PRIOR FILING DATE: 2000-11-15
; PRIOR APPLICATION NUMBER: US 09/446,504
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: PCT/JP98/02845
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: JP 9-187496
; PRIOR FILING DATE: 1997-06-26
; PRIOR APPLICATION NUMBER: JP 9-320692
; PRIOR FILING DATE: 1997-11-27
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 80
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-09-712-266-80

Alignment Scores:
Pred. No.: 1,27e-17 Length: 335
Score: 374.50 Matches: 108
Percent Similarity: 43.98% Conservative: 60
Best Local Similarity: 28.27% Mismatches: 121
Query Match: 7.01% Indels: 93
DB: 4 Gaps: 16

US-09-502-945-2 (1-2885) x US-09-712-266-80 (1-335)
QY 1547 CACCTGAGATGCTGGCCGGATCCAGAGATGCTGGCCGGCGGCGAGAGAGAGCGCTG 1606
DB 20 HisProGluAsnProLysArgLeuGluIleValLeuSerLysValArgGluLeuGlyLeu 39
QY 1607 CTAGCAAGTGCAGCGAGATCCAGATCCGACCAAGCCAGCTAGATGAGATCCAGAGCTG 1666
DB 40 GluGluArgIleGluGluProAsnProValGluGluThrPhe-----ValGluLysIle 57
QY 1667 CACTGTGAATACCAACCCCTCTCTATGGACAGTCCCTCAACCGGCAAGAGTAGAC 1726
DB 58 HisAspArgAspTyr-----IleAsnPheValLysGluAla 69
QY 1727 AGCAAGAAGTGTGGCTGCCATCAGCCAGAAAGATGATATGCTGTGGCTGTGGGGG 1786
DB 70 ValGluLysGlyIleThrArgLeuAspProAspThrTyrVal----- 83
QY 1787 ATCGGGGTGGACAGTACACCGGTGGAGATGAGATGCATCTCCAGTGTGGCGATG 1846
DB 84 -----SerProGlyThrTrp-----SerAlaIleLeuLeu 93

```

```

QY 1847 GCAGTGGGCTGCTGCTGAGCTGAGCTTCAAGTGGCTGAGAGACCTCAAGATGCA 1906
    |||
    |||
    |||
Db 94 AlaLeuGlyAla-----AlaArgSerAlaAlaLeuSerAlaLeuHisTyrGly 109
QY 1907 -----TTTGCATATCCGGCCCGACAGACACCGCCGAGAA----- 1945
    |||
    |||
    |||
Db 110 GlyLeuHisMetAlaLeuValArgProGlnHisAlaGlyArgArgGlyArgAla 129
QY 1946 -----TCCACAGCATGCTGCTTCAACTCTGATGAGCATCAACCGCAAAACTC 1999
    |||
    |||
    |||
Db 130 MetClyAlaProThrLeuGlyPheCysTlePheHisAsnAlaAlaSerAlaValThr 149
QY 2000 CTACAGCAGAGATTGAACGTGGGCAAGCTCTCATCTGAGCTGAGCATTCACCATGCG 2059
    |||
    |||
    |||
Db 150 LeuYsgLuglu---GlyValGlyLysValValValIleAspPheAspAlaHisHisGly 168
QY 2060 AATGGCACCCACAGAGCGCTTACAAATGACCCCTCTGCTACATCTCTGCATGCG 2119
    |||
    |||
    |||
Db 169 AsnGlyThrGlnGluIlePheThrAsnAspProAspValIleHisIleAspLeuHisGly 188
QY 2120 TATGACAAAGGGAAGCTTCTTCCAGGCTCTGCGGCTCTGAAAGGTTGGTGAGAGACA 2179
    |||
    |||
    |||
Db 189 ArgAsp-----LetyrProGlySerGlyAspValSerGluValGlyLysAsn 205
QY 2180 GGCCTGGGGTACAAATGTAAGCTGAGCATGACAGAGAGGTGAGACCCCGCAT- 2233
    |||
    |||
    |||
Db 206 AlaTyrGlySerIleAsnLeu-----PrometProHis 217
QY 2234 -----GGAGACGTGAGTACCTTACAGCCTTACAGGACATGTGTGATGCCATTCGCCAC 2287
    |||
    |||
    |||
Db 218 TyrSerGlyAspGlyAspTyrIleTyrValThrAspGlnIleValLeuProIleValGlu 237
QY 2288 GAGTTTCACCTGATNGTGTCTAGTCTCCGCGGTTGATGCTGCTGAGAGACAT- 2344
    |||
    |||
    |||
Db 238 GluValLysProLysValIleValIleSerAlaGlyPheAspGlyPheLysGlyAspGly 257
QY 2345 CTGTCTCTGCTGGGTGGCTACTCTGTACCCGACAGATGTTTGGCCATTCACAGAGCAG 2404
    |||
    |||
    |||
Db 258 LeuThrThrIleu-----ArgLeuThrGluSer 266
QY 2405 CTGATGACCTGAGCGAGGGGCGCG-----GTGGTGGTGGCCCTTGAGAGCA 2449
    |||
    |||
    |||
Db 267 PheTyrSerTyrAlaGlyAlaThrLeuAsnLysTyrProLeuAlaPheIleLeuGluGly 286
QY 2450 GGCCATGACTG-----ACCGCATCTGTGATGCTGCAACT 2488
    |||
    |||
    |||
Db 287 GlyTyrGlyValGlyLeuAspLysGlyPheProAlaPheIleMetGlyTyrGluGluGly 306
QY 2489 TGTGTCTGGGCTGTCTCAGTGTAAAGCTGACGCCCTTGATGAGAGCAGTCTTGACAGCA 2548
    |||
    |||
    |||
Db 307 -----LysAlaLysAlaArgGlu 312
QY 2549 AAGCCCAACATCAAGCAGTGGCCAGCGCTAGAGAAATCATGAGATCCAGCAAAACAC 2608
    |||
    |||
    |||
Db 313 GluProArgTyrGluThrLeuLysLeuAlaGluValLysAspIleLeuSerProThr 332
QY 2609 TGGAGC 2614
    |||
    |||
    |||
Db 333 TrpSer 334

```

```

RESULT 3
US-08-528-255A-1
; Sequence 1, Application US/08528255A
; Patent No. 5658016
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: FURUKAWA, YOICHI
; TITLE OF INVENTION: RPD, PROTEIN AND DNA
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLYNN, THIEL, BOUTWELL & TANIS, P.C.
; STREET: 2026 Rambling Road

```

```

; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49008-1699
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage
; COMPUTER: IBM PC/XT/AT Compatible
; OPERATING SYSTEM: MS-DOS 5.0
; SOFTWARE: WordPerfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/528,255A
; FILING DATE: September 14, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP6-227876
; FILING DATE: 22-SEPTEMBER-1994
; APPLICATION NUMBER: JP7-183763
; FILING DATE: 20-JULY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Terrance F. Chapman
; REGISTRATION NUMBER: 32549
; REFERENCE/DOCKET NUMBER: Furuya Case 1335
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (616) 381-1156
; TELEFAX: (616) 381-5465
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; LIBRARY: human fetal lung cDNA library
US-08-528-255A-1
Alignment Scores:
Pred. No.: 1.07e-12 Length: 482
Score: 302.00 Matches: 100
Percent Similarity: 42.66% Conserved: 51
Best Local Similarity: 28.25% Mismatches: 133
Query Match: 5.65% Indels: 70
DB: 1 gaps: 16
US-09-502-945-2 (1-2885) x US-08-528-255A-1 (1-482)
QY 1547 CACCTGAGCATGCTGGCGGATCCAGACATGTGTCGCCGCTGACAGACAGGCTG 1606
    |||
    |||
    |||
Db 28 HisProMetLysProHisArgIleArgMetThrHisAsnLeuLeuAsnTyrGlyLeu 47
QY 1607 CTAGCAAGTCCGAGCGGATCCAGAGTCCGAAGCCACAGCTAGATGATCAGACACTG 1666
    |||
    |||
    |||
Db 48 TyrArgLysMetGlnIleTyrArgProHisLysAlaAsnAlaGluGluMetThrLysTyr 67
QY 1667 CACTCT---GAATACACACCTCTGCTATGGACACTCCCTCAACCGCCAGAAAGCTA 1723
    |||
    |||
    |||
Db 68 HisSerAspAspTyrIleLysPheLeuArgSerIleArgProAspAspMetSerClyLys 87
QY 1724 GACAGCAAGAGTGTG-----CTCGGT-----CCATCAGCCAGCAAGATG 1762
    |||
    |||
    |||
Db 88 ---SerLysGlnMetGlnArgPheAsnValGlyGluAspCysProValPheAspGlyLeu 106
QY 1763 -----TATGCTGTGCTGCTTGTGGGGCATTCGGGGTGACACTGACACCGTGGAAAT 1816
    |||
    |||
    |||
Db 107 PheGluPheCysGlnLeuSerThrGlyGly---SerValAlaSerAlaValLysLeuAsn 125
QY 1817 GAGATGACATCTCCAGAGCTGTGCGCATGACATGAGGCTGCTCTGAGAGCTGCGCTTC 1876
    |||
    |||
    |||
Db 126 LysGlnGlnThrAspIleAlaValAsnTrpAlaGlyGlyLeu----- 139
QY 1877 AAGGTGCTGACAGAGACTCAAGAATGATTTCCATCATCCGCGCCCGACAGACACAC 1936

```





```

Db 337 ProGUAsPlyseProvaIasngIuTyTygIuTyPheGlyProAspTyThIeu 356
QY 2549 -----AGCCCAATCAACCGAGTGGCCAGTGAAGAAATCTAGATCCAG 2599
Db 357 HIsValAlaProSerAsmEtGIsuAsnLysAsnThrArgGlnGlnLeuAspSPlleaGy 376
QY 2600 AGCAAA 2605
Db 377 SerLys 378

RESULT 6
US-09-282-305-2
; Sequence 2, Application US/09282305
; Patent No. 6287843
; GENERAL INFORMATION:
; APPLICANT: Baldwin, Donald A.
; APPLICANT: Baldwin, Steven P.
; APPLICANT: Crane, Virginia C.
; TITLE OF INVENTION: Maize Histone Deacetylases And Their Uses
; FILE REFERENCE: 5718-44,
; CURRENT APPLICATION NUMBER: US/09/282,305
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/080,563
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 2
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Zea mays
US-09-282-305-2

```

```

Alignment Scores:
Pred. No.: 4,7e-11 Length: 458
Score: 277.50 Matches: 99
Percent Similarity: 44.00% Conservative: 66
Best Local Similarity: 26.40% Mismatches: 143
Query Match: 5.19% Indels: 68
DB: 4 Gaps: 14
US-09-502-945-2 (1-2885) x US-09-282-305-2 (1-458)

```

```

QY 1400 GCSCTGGGCGGAGCCAAATCTCCCTGCTGCGCCCTGGGGGATGAAGAAGCCCGGAGAC 1459
Db 3 AlaSerGlyGlyGlyAlaSerLeuPro---SerProAlaGlyGlyGluAspLanHisArg 21
QY 1460 CAACCCGTAACACACCTCTTCAACACAACTGGTGTAGACACAGTTCATGCTAAAGCAC 1519
Db 22 ArgAlaGValSerTyrrPheTyrrGluProSerIle----- 32
QY 1520 CAGTGCATGTGGCGGAGACACACAGTGT-----CACCTGAGCATGTGGCCGATG 1570
Db 33 -----GlyAspTyrrTyrrGlyGlnGlnHisProMetLysProHisArgIle 48
QY 1571 CAGACCATGTGCTCCGCTGAGAGACAGAGCTGCTTAAAGTAAGTGGAGCGGATGCGA 1630
Db 49 ArgMetAlaHisSerLeuValAlaHisTyrrGlyLeuHisArgLeuLeuGluLeuSerArg 68
QY 1631 GGTTCGAAAGCCAGCTAGATGATCCAGACAGTGCATGTAATTAACACACCCCTGTC 1690
Db 69 ProTyrrProAlaSerGlnAlaAspIleArgArgPheHisSerAspTyrrValAlaPhe 88
QY 1691 TATGGAGACAGTCCCTCAACCGGAGAGATGACAGAGAGA--GTTGCTGCGTCCCAT 1749
Db 89 LeuAlaSerAlaThrGlyAsnProGlyValLeuAspProAlaAlaIleLysArgPheAsn 108
QY 1750 CAGCAGAAAGATGTATGCTGTGCTGCTGCTGGGGGATCGGAGGTGACAGTGAACCGT 1809
Db 109 ValGlyGluAsp-----CysProValPhe----- 116
QY 1810 GTGGATAGATGACACTGCTCCAGTCTGTGCGCATGGAGTGGGCTGCTGAGACT 1869
Db 117 -----AspGlyLeuPheProPheCysGlnAlaSerAlaGlyGlySerIleGlyAla 133

```

```

QY 1870 GGCCTTCAAGTGTGCTGACAGAGAGCTCAAGATGGATTGGCATATCCGGCCCCAGG 1929
Db 134 -AlaValLysLeuAsnArgGlyAspAlaAsp-----IleThrValAsnThrAlaGln 150
QY 1930 A-----CACACCGCCGAGGAATCCACAGACCATGGGATTTGCTTCTTAAGTCTGTAGC 1983
Db 150 yGlyLeuHisHisAlaLysLysSerGlnAlaSerGlyPheCysTyrrValAlaAspSPlleaV 170
QY 1984 CATCACCCGAAAGACCTGTACAGACAGTGAACGTGGGAGAGCTGCTCATGCTGAGACTG 2043
Db 170 IleuAla-----IleLeuGlnLeuLeuLysPheHisArgArgValLeuTyrrValAspIle 188
QY 2044 GCACATTCACATGGCATGGCAATGGACCCAGGAGCGTTTCAATGAACCCCTGTGCTTA 2103
Db 188 eAspValHisHisGlyAspGlyValGlnGlnAlaPhePheThrThrAsnArgValMetTh 208
QY 2104 CATCTCTGTGCATCCGCTATGACACAGGAACTTTTCCAGGCTCTGGGGCTCCGTGAAGA 2163
Db 208 rValSerPheHisLysTyrr-----GlyAspPhePheProGlyThrGlyHisIleThrAs 226
QY 2164 GGTGTGGAGAGACAGCGGTGGGTACAAATGAAACGTGGCATGGAGACAGAGGTGTGA 2223
Db 226 rValGlyAlaAlaGlyGlyLysHisTyrrAlaLeuAsnValProLeuSerAspGlyIleAs 246
QY 2224 CCCCCCATTTGGAGACAGTGGAGTACCTTACAGCCTTTCAGACAGACAGTGTATGCCATTGC 2283
Db 246 P-----AspThrThrPheArgGlyLeuPheGlnCysIleIleLysLysValMet 262
QY 2284 CCACAGTTCTACCTGATGTGCTCTACTCTCCCGGGGTTGATGCTTGAAGAGACA 2343
Db 262 tGlnValTyrrGlnProAspValAlaValLeuGlnCysGlyAlaAspSerLeuAlaGly-- 281
QY 2344 TCTGTCTCTGCTGGGTGCTACTGTGTGCACCCGAGATGTTTGGCCAC----- 2392
Db 282 ----AspArgLeuGlyCysPheAsnLeuSerValLys-----GlyHisAlaAspCysLe 298
QY 2393 -----TTGACACGAGCAGTGTATGACCCCTGAGGAGGGCGGGGTGG---- 2432
Db 298 uArgPheLeuArgSerTyrrAsnValProMetMetValLeuGlnGlyGly--GlyTyrrThrI 318
QY 2433 -----TGTGCGCCCTGGAGGAGGCCATG 2456
Db 318 leArgAsnValAlaArgCysTyrrCysTyrrGlnThrAlaVal 331

RESULT 7
US-09-548-372D-13
; Sequence 13, Application US/09548372D
; Patent No. 6420534
; GENERAL INFORMATION:
; APPLICANT: GURNEY ET AL.
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USE
; FILE REFERENCE: 29915/62801
; CURRENT APPLICATION NUMBER: US/09/548, 372D
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: US 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 2088
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-548-372D-13

```

Alignment Scores:

Pred. No.:	1e-10	Length:	2088
Score:	275.50	Matches:	191
Percent Similarity:	31.12%	Conservative:	58
Best local Similarity:	23.88%	Mismatches:	336
Query Match:	5.11%	Indels:	215
DB:	4	Gaps:	37

US-09-502-945-2 (1-2885) x US-09-548-372D-13 (1-2088)

QY	2105	TGTAGACACAGAGGGGCTCTTGTATGAAGCCGTGGCGGGGTGCATTGGCATATGGTGAATCT	2046
Db	82	CysThrGlyCysThrGlyCys---ThrGlyAlaAlaCysCysCys-----Cys	97
QY	2045	CCGAGTCCACGATGAGACGCTTGGCCACGCTCACTTCTGCTGTAGA-----	1998
Db	98	AlaGlyAlaThrThrThrGlyCysCysAlaThrInGlyThrCysThrGlyYThrGlyCys	117
QY	1997	-----GTTTGGCGGTATGGCTTCAAGATTTGAAGAAGCAAGAAATC	1959
Db	118	AlaGlyAlaCysThrGlyAlaAla-CysAlaThrGlyCysAlaCysAlaThrGlyAlaAl	137
QY	1958	CCATGCGCTGTGATATCTCTCGCGCTGGCTGTGCTCCGGGGCCGATGATGGCAATTCATCT	1899
Db	137	aThrGlyYThrCysCysAlaGlyAlaAlaThrGlyGlyGlyAlaAlaGlyYThrGlyGly	157
QY	1898	TGAAGCTCTCTGAGCCACCTTGAAGCCAGCTCCAGCAGGACGCCACTGGCATGCGCA	1839
Db	157	YAlaThrThrCysAla-----GlyAlaThrCysCysAlaTh	169
QY	1838	CAGACACGAGAGAGTGCATCTTCCACACGCTGCATCTGCATCCACCCGATGCCCCAC	1779
Db	169	rCysAlaGlyGly-GlyAlaCysCysAlaAlaAlaAlaCysCysThrGlyCys-----	186
QY	1778	AAGCAGACACAGCATCATCTTCTGCTGATGAGGACCGACCACTTCTGCTGCTACT	1719
Db	187	-----AlaThrThrGly-----AlaThrAlaCysCysAlaAla	197
QY	1718	TCGTCCGGTTAGAGGCGACTGTGCTCCATPAGACAGGGTGTGT--ATTTCAGATGCACTG	1662
Db	198	GlyGlyAlaAlaGly-----GlyCysAlaThrCysCysThrGlyCys	211
QY	1661	TCTGATATCATCTAGCGTGGCTTTGCGACTCGCATCCGCTCGGCACTTC-----	1611
Db	212	AlaGlyThrAlaThrThrGlyCysCysAlaAlaGlyAlaAlaGlyYThrCysThrAlaCys	231
QY	1610	-----TAAAGCAGCGCTGTCTCTGACGCGGAGCAGATGC-----	1575
Db	232	CysCysThrGlyAlaAlaAlaCysThrGlyCysAlaGlyAlaThrCysAlaCysCysAlaAla	251
QY	1574	TCTGATCCGCGCCAGCATGCTCTCAGGGGTGCACGTGTGTTCCCGACATGC-----	1524
Db	252	ThrGlyThrGlyGlyYThrAlaGlyAlaAlaGlyCysCysAlaAlaCysCysAlaAlaCys	271
QY	1523	-----ACTGGTCTTTAGCATGAACGTGTCTGACACCACTTGTGTGAAGAGT	1473
Db	272	CysAlaGlyThrGlyAlaCysCysAlaThrCys-----Cys	283
QY	1472	GCTTGAAGGCTTGTGTGCGGGCTTCTTATGATCCGCCAGGCGACAGAGGAGA-----	1418
Db	284	AlaGlyAlaAlaCysThrGlyGlyYThrGlyCys-AlaAlaGlyCysGlyGlyGlyGlyCys	303
QY	1417	TTGGGTAGCGCCAGGCGTTGTGGTGGGACAGATGGCCAGCGTGAAGGGCGCT-----	1366
Db	303	scYsglyCysAlaAlaGlyCysAlaGly---ThrGlyCysAla-AlaGlyAlaCysCysC	322
QY	1365	-----T	1365
Db	322	ysAlaThrCysCysCysAlaCysThrThrThrGlyYAlaThrThrThrCysCysC	342
QY	1364	GGTACACCTGCACAAAGTTGACAGGGGTGGGCACTGTGAACAAGTTTTTTGTATCCAGAC	1305
Db	342	ysThrAlaCysCysGlyCysThrGlyCysThrThr-----	353

[illegible]





Db 446 lacys---ThrGlyGlyCysAlaCysAlaCysCysGlyThrCysGlyCysCysAlaAla 465  
QY 953 GGGGACTGCGCGGCGGAGGAGTGC---GCTCAGAGGCGCATGGCGCGGA 897  
Db 465 laGly---AlaGly-AlaCysAlaThrGlyCysAlaGlyThrGlyAlaAlaGly 483  
QY 896 GCT-----TGCTACCGTCGCTGCTG 870  
Db 484 AlaGlyThrAlaCysAlaAlaCysThrThrGlyCysAlaThrGlyAlaCys----- 501  
QY 869 CCACAGCTTCACCCGTCAGTGGGAGTCCCGTCGACGTCACAGCAATAGGTCG 810  
Db 502 -----ThrAlaCysGlyGlyCysAlaThrThrGlyThrGlyCysThrGly 517  
QY 809 TCTGTCGCGG-----GGCTGCTCCAGCAAGCAAT 777  
Db 518 ---CysCysThrGlyCysGlyGlyAlaAlaThrThrGlyAlaCysAlaAlaGlyThrC 537  
QY 776 GCTGCAGAGGAGGAGCATGCCGTCGCGCTCCGCTCCAGTCAGTCGCCAGCA 717  
Db 537 yscysGlyAlaGly-----GlyGlyThrAlaGlyAlaGlyThrThrGly 554  
QY 716 GGCAGCCAGGAATAGAGATGCTCATGACTTGCCTGACGTCGCCCTGCCGA 657  
Db 554 hrcGlyThrGlyThrGly---CysCysCysAlaCysThrGlyGlyCysThrGlyAlaAla 573  
QY 656 -----GGGACTGAGGCGCTGCTGCTGCG 633  
Db 574 GlyAlaAlaAlaGlyThrGlyAlaCysAlaAlaThrGlyThrGlyAlaThrCys 593  
QY 632 CCTCCTGCTGTG---TCGACAGCTCGGAGGAGCAGTGAAGTGAAGTGAAGTGA 576  
Db 594 ThrGlyCysThrGlyAlaThrGlyCysGlyGlyAlaGlyAlaGly----- 609  
QY 575 CCGTGGCTGCAGCCCTAGG-----AGATGTTGGCAGAGAGAGAGCTGTAGAGC 522  
Db 610 -----GlyAlaThrGlyAlaCysThrCysGlyGlyAlaThrGlyThrCysThrGly 626  
QY 521 TGAACGTGTTGGGAGGAGTCCAGAGGAGG----- 499  
Db 627 GlyThrGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 646  
QY 488 -----CTCGTGTGAGGAGCATCTCAGTGGGAGTGGGAGTGGAGCCAG 441  
Db 647 AlaCysThrAlaThrGlyCysAlaGlyAlaThr-GlyGlyGlyAlaGlyThrGlyAlaAl 666  
QY 440 TAAAGCCATTCGACGAGTGTGCTGTGGAGCTGTGGAGAGCTGGGCGGAGCCGG 381  
Db 666 aGlyAlaCysAlaAlaAlaGlyThrAlaGlyThrAlaGlyAlaAlaGlyThr---AlaGly 685  
QY 380 GTGCGCTTTACACAGGAGGAGCCCGCCAGCCGCTGATCTCAACAGCTCT 321  
Db 685 yCysAlaGlyAlaGlyGlyAlaGlyGlyAlaAlaGlyAlaGlyThrGlyGly----- 703  
QY 320 TCTTAAAGTGTAAATACAGTCCATCTTGGCAGCAGAGGAGGAGCTTCTCGCT 261  
Db 704 -CysThrGlyAlaGlyGlyThrGlyAlaAlaGlyAlaAlaGlyAlaAla----- 720  
QY 260 CAGCCACCTTCTGTTTACCTTGAACGCACTTCAAGTTGGTTCAGAGCTGTTTGC 201  
Db 721 -----GlyAlaAlaGlyGly---Cys 726  
QY 200 GGAAGG---GGAAGTGTCTCGACTGCTGTAAGGCGCCAGCAAGGAGTGTAGAGG 144  
Db 727 GlyAlaThrGlyAlaThrGlyAlaCys-----GlyAlaGlyAlaCysGlyAla 743  
QY 143 GAGGCTCCAGGAGGCGCTGCGGAGAGGAGCAATCTGCCAAGAAGACATGGTG 84  
Db 744 ThrGlyAlaGlyAlaThrGlyGlyThrGlyAlaThrGlyAla-----GlyGly 760  
QY 83 CTCGCCAGCATTTGGGCTGCTGGAGGAGGAATGT 48  
Db 761 ThrAlaGlyAlaGlyAlaAlaGlyAlaGly 772

RESULT 9  
US-09-282-305-6  
Sequence 6, Application US/09282305  
Patent No. 6287843  
GENERAL INFORMATION:  
APPLICANT: Baldwin, Donald A.  
APPLICANT: Briggs, Steven P.  
APPLICANT: Crane, Virginia C.  
TITLE OF INVENTION: Maize Histone Deacetylases And Their Uses  
FILE REFERENCE: 5718-44,  
CURRENT APPLICATION NUMBER: US/09/282,305  
CURRENT FILING DATE: 1999-03-31  
PRIOR APPLICATION NUMBER: 60/080,563  
PRIOR FILING DATE: 1998-04-03  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 6  
LENGTH: 439  
TYPE: PRF  
ORGANISM: Zea mays  
US-09-282-305-6

Alignment Scores:  
Pred. No.: 1,18e-10 Length: 439  
Score: 271.50 Matches: 90  
Percent Similarity: 42.94% Conservative: 59  
Best Local Similarity: 25.94% Mismatches: 107  
Query Match: 5.08% Indels: 91  
Gaps: 16

US-09-502-945-2 (1-2885) x US-09-282-305-6 (1-439)

QY 1547 CACCCATGAGATGCTGGCCGATCCAGAGATCTGGTCCGCGCAGAGAGCTG 1606  
Db 41 HisProMetLysProHisArgIleArgMetThrHisAlaLeuGlyArgTyrGlyLeu 60  
QY 1607 CTTAGCAATGCGAGCGGATCCGAGTCCGAGCCAGCCAGCTAGATGATCCAGACAGT 1666  
Db 61 LeuAspIleMetGlnValPheArgProHisProAlaArgAspArgLeuCysAlaPhe 80  
QY 1667 CACTCT---GATATCACACACCTCTCTATGGAGCAGTCCCTCAACCGGAGAACTA 1723  
Db 81 HisAlaAspArgTyrValSerPheLeuArgSerValThrProGluThrGlnGln----- 98  
QY 1724 GACAGCAGAGATTGCTCGT-----CCATCAGCCAGAG 1759  
Db 99 AspGlnIleArgAlaLeuLysArgPheAsnValGlyLeuAspCysProValPheAspGly 118  
QY 1760 ATGATGCTGTGCTGCTGCT-----GGGGGATCGGGGAGGAGTGAACACGCTG 1810  
Db 119 LeuTyrSerPheCysGlnThrTyrAlaGlyLysValIly----- 132  
QY 1811 TGGATATGATATGCACTCTCTCAGTGTGTCGATGAGCAGTGGAGCTGCTGAGACTG 1870  
Db 133 -----GlyAlaValLysLeuAsnHisGly-----HisAspIle 143  
QY 1871 GCCTTCAAGTGGCTGCAGAGAGCTCAGATGATTTCCATCCGCGCCCAAGGA 1930  
Db 144 AlaIleAsnThrPheAlaGlyLeu----- 151  
QY 1931 CACCGCCGAGGAGATCCAGCATGAGTGGATTTGCTTTCACATCTGTAAGCATCAC 1990  
Db 152 HisHisAlaLysLysCysGlyAlaSerGlyPheCysTyrValAsnAspIleValAla 171  
QY 1991 GCAAACTCCTACAGCAGAGATTGAACGTGAGCTGGCAGGCTCTCATCTGAGTGGACAT 2050  
Db 172 -----IleLeuIleuLeuLeuLysTyrHisGlnArgValLeuTyrAlaAspIle 189  
QY 2051 CACCATGAGCATGAGCAGCCAGAGCGTGTCTACATGACCCCTGTGCTATCATCTGT 2110  
Db 190 HisHisGlyAspLeuValAlaGlnGluAlaPheTyrThrAspArgValMetThrValSer 209

```

OY 2111 CTCATCGCTATAGACACGGAACTCTTTCCAGGCTGTGGGCTCTGAGAGGTTGGT 2170
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 210 PHEHSLYSRPE-----GLYASPTYRPHPEPROGLYTHRGLYASPLAARGSPVALIGLY 227
OY 2171 GGAGGACCGGCTGGGCTGCATATGTAAGTGCATGAGCAGAGGTGTGACCCCCC 2230
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 228 HHSERLYSGLYLSTYRYSERLEUSNAVALPROLEAUSAPSPGLYLEASP----- 245
OY 2231 ATGGAGCGTGGAGCTTACAGCTTCAGGACGACAGAGGTGATGCCCCCAGCAG 2290
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 246 -----ASPLGUSERTYRGINSERLEUPHELYSPROILEMELGLYLVSVALMETGLI 263
OY 2291 TTTCACCTGATGTGCTCTACTCTCCCGGGGTTTGTAT----- 2329
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 264 PHEASNPROGLYALVALLEUGLNCYSGLYALASPSERLSEUSERGLYASPARGL 283
OY 2330 -----GCTGTGAAGACATCTGTCT----- 2350
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 284 GLYCYSRPEASNLSEUSERTILEYSGLYHISALAGLUCYASVALARGPHEMETARGSERPHE 303
OY 2351 -----CCTCTG-----GGTGGCTACTCTGTCT-----ACGCCAGATGT 2383
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 304 ASNALVALPROLEULLEULLEULGLYGLYGLYTYRTHRILEARGASVALALARGCY 323
OY 2384 TTTCGCCACTTGACCGAGGACGCTGATGACCCCTGGCAGGGGCGGTGCTGGCCCTG 2443
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 324 TPCCYSTYRGLUTHRGLY-----VALALALEU 332
OY 2444 GAGGAGCGCATGATCTGACC 2464
Db 333 -----GLYHISGLULLEUTHR 337

RESULT 10
US-09-282-305-4
: Sequence 4, Application US/09282305
: Patent No. 6287843
: GENERAL INFORMATION:
: APPLICANT: Baldwin, Donald A.
: APPLICANT: Brilggs, Steven P.
: APPLICANT: Crane, Virginia C.
: TITLE OF INVENTION: Maize Histone Deacetylases And Their Uses
: FILE REFERENCE: 5718-44,
: CURRENT APPLICATION NUMBER: US/09/282,305
: CURRENT FILING DATE: 1999-03-31
: PRIOR APPLICATION NUMBER: 60/080,563
: PRIOR FILING DATE: 1998-04-03
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 4
: LENGTH: 351
: TYPE: PRT
: ORGANISM: Zea mays
US-09-282-305-4

Alignment Scores:
Pred. No.: 1,28e-10 Length: 351
Score: 270.50 Matches: 69
Percent Similarity: 49.14% Conservative: 45
Best Local Similarity: 29.74% Mismatches: 77
Query Match: 5.06% Indels: 41
DB: 4 Gaps: 8

US-09-502-945-2 (1-2885) x US-09-282-305-4 (1-351)
OY 1846 GGCACGTGGGCTGCGTGGAGCTGGCCCTCAAGTGGCTGCAGAGAGAGCTCAAGATCG 1905
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 7 GLYALASERLEUPROSETRALGLYGLIUSAPRALANISARGARGARGVALSERLYR 26
OY 1906 ATTTCSCAT-----CATCCGGCCCCCAGAGACA 1932
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 27 PHELYRGLUPROSERILEGLYASPTYRITYRGLYGLINGLYHISPROMET-LYSPROHI 46
OY 1933 CCACGCCGAGAAATCCACAGCCATGGGATTCGTCTTCAACTCTGATGCCATCAGCCG 1992

```

```

Db 46 SHISALALYSSESRGLUALASERGLYPHECYSYRVALAASAPRIILEVALALEA-- 65
OY 1993 AAAACTCTACAGCAGAACTTGAAGCTGGGCAAGTCCCTCATCGTGGACATCA 2052
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 66 ---ILEUGLILEULLEULYSRPHENISARGARGVALLEUYRVALIAPRIILEASVALHI 84
OY 2053 CCATGGCAATGACACCCAGCAGGCTTCTACAAATACCCCTGTGCTTCAATCTCTCT 2112
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84 SHISGLYASPGLYVALIGLUALALPHEPHERTHRASNALARGVALMETTHRVALSERPH 104
OY 2113 GCATCGCTATGACAAAGGCACTTTTCACAGCTGTGGGCTCTGAAAGAGTTGGTG 2172
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 104 EHISLYSTRY-----GLYASRPHPEPROGLYTHRGLYHISILERTHRASPVALIGLY 122
OY 2173 AGGACACGGCGTGGGTTCATGATGACGTGGCATGAGCAGAGAGGTGGTGGCCCCCAT 2232
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 122 AALAGLUGLYSHISTRVALALALEUASNAVALPROLEUSERTASRGLYLEASP----- 139
OY 2233 TGGAGACGTGAGATACCTTACAGCCTTCAGACAGTGTGATGCCATTTGCCAGAGTT 2292
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 140 ---ASPTHRTHRPHEARGLYLEUPHEGLNCSILERTILEYSLYSVALMETGLI 158
OY 2293 CTCACCTGATGTGCTTCATGTCCTCCGCGGTTTATGCTGTGAAAGACATCTCTCC 2352
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 158 TGLNPROSPVALIIVALILEUGLNCYSGLYALASPSERLSEUALAGLY-----ASPAR 176
OY 2353 TCTGGTGCTGCTCTGTGACCGCCAGCGAGATGTTTGCCAC----- 2392
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 176 GLEULCYSPHENASNLSEUSERTVALYS-----GLYHISALASPCYSLEUARGPHELE 194
OY 2393 -----TTGACCAAGCAGCTGATGACCTTGACAGGGGCGGTG----- 2432
Db 194 WARGSETRYAENVALPROMETMETVALLEUGLYGLY-GLYTYRTHRILEARGASNV 214
OY 2433 -----TGCTGGCCCTGGAGGAGGCCANG 2456
Db 214 AALARGCYSTRPCYSTYRGLUTHRALALVAL 224

RESULT 11
US-08-630-915A-37
: Sequence 37, Application US/08630915A
: Patent No. 6309820
: GENERAL INFORMATION:
: APPLICANT: SPARKS, Andrew B.
: APPLICANT: HOFFMAN, No. 6309820h
: APPLICANT: KAY, Brian K.
: APPLICANT: FOWLER, Dana M.
: APPLICANT: MCCONNELL, Stephen J.
: TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
: TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
: NUMBER OF SEQUENCES: 227
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pennie & Edmonds LLP
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/630,915A
: FILING DATE: 03-APR-1996
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Mistrock, S. Leslie
: REGISTRATION NUMBER: 18,872
: REFERENCE/DOCKET NUMBER: 1101-174

```







Db	900	-----	AlaGlyThrCysCysGlyThrThrGlyThrCysCysAlaCysCys	913	
Qy	2645	-----	GGCCAGACCAAGGCGAGACTTCTGCACACAGCTCCAGTGTTCCTGTGATCT	259	
Db	914	GlyAlaGlyThrCysThrThrAlaThrGlyAlaGlyThrThrGlyCysCysGly--CysA	933		
Qy	2591	CGATACCTTCTCTAGCTGGCCACCTGCCTGATAGTGGCTTTGTCTGCAGACTGCC	253		
Db	933	laGlyThrThrThrThrAlaCysThrThrCysAlaGlyThrCysThrAlaThrCysA	953		
Qy	2531	CATCCAGAGGC--	-TGCAGCTTAACTAGACAGAGCGCGAGACACAGCTT	248	
Db	973	laAlaAlaGlyCysThrThrAlaGlyThrThrCysAlaGlyGlyAlaGlyAlaThrA	993		
Qy	2483	CAGAGCATCAACAGATGGCGGTCAAGTCATGGCTCCCTCCAGGCGCAGCACCCGGC	242		
Db	2423	CCCCGCCAAGGTCATCACTGCCCTGGTCAAGTGGCCAAAACATCTGGCGGTACAGACT	236		
Qy	993	laThrThrCysCysGlyCys--	-----	100-	
Db	2363	AGCCACCAAGAGAGACAGAGTCTCTCAACAGCATCAAAACCGGCGAGACTAGAGACA	230		
Qy	1001	-----	-CysThrThrThrGlyCysThrGlyGlyThrThrCysThrT	101	
Db	2303	CA---TCAGGTGAGAACTCGGGCCAAATGGGCATCACCACTGTCTGMAAGCTGAAGCT	224		
Qy	1014	hrCysCysThrAlaThrThrCysAlaAlaGlyCysGlyThrAla--	-----	102-	
Db	2246	ACTCCAGCTCTCCAAATGGGGGGCTCCACA--	-----	221-	
Qy	10239	--AlaCysAlaThrCysGlyCysThrThrCysThrAlaCysAlaThrCysAlaThrC	1044		
Db	2217	-----	CCCTCGTCCATGCGCCAGCTTCAATTTGATGCCAGGCTGGTCTCCACCAACT	216	
Qy	1048	ysCysAlaAlaAlaThrAlaThrAlaThr--CysThr--	-----	105-	
Db	2162	CTTTCAGGAGCCCCAGAGCTTGAAGAAGTTCGCCGTTCATAGCGATGCAGAGAGATGT	210		
Qy	1060	-----	AlaCysAlaThrCysThrGlyThrCys	106	
Db	2102	-----	AGACACAGAGGGGCTCATTTGAGAACGCGCTGGGTGCATTC--	205	
Qy	1070	CysThrGlyAlaThrThrGlyThrCysGlyThrCysAlaThrCysThrAlaThrGly	108		
Db	2057	-----	CATGGTGAATGTCCACAGTCCACGATGAGAGACTTCCACGTTCACTTGTCT	200	
Qy	1090	GlyGlyThrAlaThrCysThrThrThrThrThrGly--	-----	GlyThrThrThrGly	110-
Db	2003	GTAGAGATTTTGGCGTGTGCTACAGATTGAAGAAGACAGAAATCCCATGGCTGGATGT	194		
Qy	1106	AlaGlyCys--CysGlyCysGly--	-----	AlaThrGlyCysGly-A	111
Db	1943	CCTGGCGGTGTGTCTGGGGGGCGGATGATGGCAAAATCCATTGTGAGCTCTCTGCAG	188		
Qy	1118	laCysAlaCysGlyGlyGlyCysAlaAlaGlyAla--	-----	GlyThrThrGlyC	113
Db	1883	CCACCTTAAAGCCAGCTCCACGAGCGAGCCCACTGCCATGGCCACAGCACTGGAGAGT	182		
Qy	1134	ysThrCysThrCysGlyGly--AlaGlyThrCysAlaCysCys--	-----	-----	114
Db	1823	GCATCTCATTCACAGGCTGCTACCTGCA--	-----	-----	179
Qy	1148	AlaCysThrGlyThrGlyCysThrCysAlaCysAlaAlaThrGlyAlaCysThrAlaCys	116		
Db	1793	-----	-----	-----	175
Qy	1168	ThrThrThrGlyAlaThrGlyAlaCysCysAlaThrGlyAlaCysThrAlaAlaThrAla	118		
Db	1754	GGCTGAATGGAGCAGCAACTTCTTGCTGTCTAGCTTCTGCGGTGAGGGAGCTGGCC	169		

D	b	1188	GlyThrThrCysAlaIleThnGlyCysCysAlaIleAlaIleAlaIleGlyThrGlyThrCysAla	1207
Q	y	1694	CATAGACAGGGGTGGTATTACAGAGTCAGCTGTGGATCTCATCTAACCAGTGCCTTGC	1635
D	b	1208	ThrAlaThrGlyThrGlyAlaAlaAlaAla-----	1217
Q	y	1634	GACCGGATCGGCGCACGTGTGTAAAGCAGGGCTGTCCTGCACAGCGGACAGATGC	1575
D	b	1218	----GlyThrAla---Thrcys-----GlyAlaThrAlaThrAlaThrThrcys	1233
Q	y	1574	TCTGGATCCGAGCA-----GCATGCTCAGGGGTGCAGCTGTGTGTCCCGACATGCAC	1519
D	b	1233	ThraIaGlyThrThnGlyThrcysThnGlyCysethr-ThrcysAlaThrGlyAlaThrGl	1255
Q	y	1518	TGCTTAGCATGAACGTGTGCTAGAACCAACTTGTGTGAAGAGTCTTGACGGTTGG	1458
D	b	1252	yGlyThrAlaThrThrcys--ThrglyThrThrcys-----	1263
Q	y	1458	TCTGGGGGTCTTCATGCCCCCAGGGGACAGGGGAGATTGGGTACGGCCAGGGCT	1399
D	b	1264	-----AlacysThrThrcysThraIaGlyAlaAlaThrAlaCysGlyC	1278
Q	y	1398	TGGTGGGCACAGTGGCCAGGCTGAGGGGCGCTGTGTACACTGCATAAGTTGACAGCGT	1333
D	b	1278	yscysGlyCyscysGlyThrcys-GlyGlyAla---ThralacysAlaThrCysAlaGly	1296
Q	y	1338	TGGGCACTGTGAACAAGTTTTTGTATCCAGCACAGGCTCTCCAG-----	1291
D	b	1297	CysAlaAlaIaIaCysGlyGlyAlaThrGlyAlaIaGlyAlaGlyThrcylThrcysCys	1316
Q	y	1290	TGCGGCCCTCCTCAGACCAACTGTGCCCTCCTGCTTAACCTGATGCATCTCC	1231
D	b	1317	GlyAlaGlyCysCysAlaGlyAlaAlaAlaIaAlaIaGlyAlaAlaThrCysThrcysGlyAla	1366
Q	y	1230	TCCCTCCGCCCATCTCTCCGTCGTCCTCCTCCAGGCTCTCCGTGCTGCACATC	1171
D	b	1337	AlaThrGlyCysThrGlyAlaIaCysCysCysThrThrThralacysAlaCysAla	1356
Q	y	1170	TCTGTGAGCCCTCCGAGGCACTGATCAGGGCTCCCTCCACAGCAAGCATCTGCTGC	1111
D	b	1357	ThrcysThrThrGlyAlaGly-----ThrcysThrcysThrThrcysThrcysCys	1374
Q	y	1110	TCCGTCACTCTCCTCCTGTCCTCTCAAGGTGGGTGGTGGCTGCCTGGGCAAGCTCCCT	1051
D	b	1375	AlaAlaAlaIaIaCysGlyThralaIaCysThrcysThrAlaThrCysCysGlyThrThr---	1393
Q	y	1050	GTCCTGTGAGAGATCTTGCCACAGCTAGAGTGC-----	1018
D	b	1394	-----CysCysCysThrcysGlyThralaThrThrThrcysAlaIa	1407
Q	y	1017	-----TCTGCTCTCTGCTTCTCCAGCAAGCTGTGCTGTGT-----	979
D	b	1408	CysAlaIaIaCysAlaCysCysAlaGlyThralaCysCysGlyCysCysCysGlyThr	1427
Q	y	979	-----	979
D	b	1428	ThrThrThrAlaCysThrcysGlyThrcysCysAlaCysCysGlyAlaThrCysAlaIa	1447
Q	y	978	-----TGGATGCCAGCTGCTGC	961
D	b	1448	AlaCysGlyThrcysCysAlaIaIaCysCysThrlYThralaCysAlaThrThrcysCys	1467
Q	y	960	AGGCGCTGGGACCTCCGCGACGCGTGAAGATGATCGGCTCAGGGCCGATGCCGC	901
D	b	1468	GlyGlyAlaGlyThrcysGlyCysAlaGlyCysGlyCysAlaCys---GlyAlaCysGly	1486
Q	y	900	GGGAGCTTGCTTACCGTCCGATCTGTGGCCACAGTTCAACCCGTACTAGTGGGAC	841
D	b	1487	AlaThrThrThrThrcysThrcysAlaAlaIaThrGlyAlaGlyGlyAlaThrGly---	1505
Q	y	840	TGCCCGTGGAGTGGCACAGCATGAGGTCCTGCTGC-----CGGCC	796
D	b	1506	CysAlaGlyThrGly-----CysCysGlyAlaAlaIaThrGlyAlaIa	1519







[illegible]

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 21, 2003, 08:38:27 ; Search time 70.4665 Seconds

(Without alignments)  
10910.938 Million cell updates/sec

Title: US-09-502-945-2

Perfect score: 5346  
Sequence: 1 ggaattcccttctgtcgaagt.....aaacaaaagttaaaattt 2885

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL-frame+ n2p model -DEV-xlp  
-Q/cgnt\_1/USPPO.spool/US09502945/runat\_14032003\_101057\_19087/app\_query.fasta.1.10979  
-DB-A.Geneseq.101002 -QEMT-fastan -SUFFIX-frag -MINMATCH=0.1 -LOOPCL=0  
-LOPEXT=0 -UNITS-bits -START=1 -END=-1 -MATRIX-blosum62 -TRANS-human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE-Pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE-LOCAL -OUTFMT=ptc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09502945 -CGCN\_1.1.365 -runat\_14032003\_101057\_19087 -NCPU=6 -ICPU=3  
-NO\_XLPYX -NO\_MMAP -LARGEOUTRY -NES\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A.Geneseq.101002:\*  
1: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/genesep-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/genesep-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/genesep-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4798	89.7	1141	21 AAB43008	Human ORFX ORF2772
2	4669	87.3	897	20 AAY07092	Colon cancer assoc
3	3659	68.4	716	22 AAB49958	Human histone deac
4	3612.5	67.6	717	23 AAO99660	Human histone deac
5	2691	50.3	967	22 AAB49957	Human histone deac
6	2685	50.2	967	23 AAO99659	Human histone deac
7	2127	39.8	417	22 AAM93526	Human histone deac
8	1831.5	34.3	1008	22 AAM78891	Human polypeptide
9	1831.5	34.3	1020	22 AAM79875	Human protein SEQ
10	1739.5	32.5	1428	21 AAY97023	Caspase 8-Interact
11	1671.5	31.3	835	21 AAY99662	Human histone deac
12	1607.5	30.1	614	22 AAB92700	Human protein sequ
13	1601.5	30.0	1235	22 AAB58493	Drosophila melanog
14	1573	29.4	574	21 AAY97032	Caspase 8-Interact
15	1552.5	29.0	541	21 AAB42845	Human ORFX ORF2609
16	1484.5	27.8	425	23 AAB89750	Human polypeptide
17	1296	24.2	328	21 AAB45473	Human secreted pro
18	1103.5	20.6	327	21 AAB45474	Human secreted pro
19	945	17.7	191	22 AAB94542	Human protein sequ
20	911.5	17.1	276	22 AAB92577	Human protein sequ
21	878.5	16.4	361	22 AAO23362	Novel human enzyme
22	771.5	14.4	1215	23 AAE22856	Human histone deac
23	767.5	14.4	1215	23 AAO99661	Human histone deac
24	741.5	13.9	726	22 AAB93058	Human protein sequ
25	733.5	13.7	883	22 AAB62141	Drosophila melanog
26	708	13.2	552	21 AAG51590	Arabidopsis thalia
27	704	13.2	499	21 AAG51591	Arabidopsis thalia
28	702	13.1	468	21 AAG51592	Arabidopsis thalia
29	700	13.1	552	21 AAG09530	Arabidopsis thalia
30	696	13.0	499	21 AAG09531	Arabidopsis thalia
31	694	13.0	468	21 AAG09532	Arabidopsis thalia
32	645.5	12.1	817	22 AAB95640	Human protein sequ
33	623.5	11.7	662	23 AAB81614	Human HDAC9 vari
34	623.5	11.7	673	23 AAB81613	Human HDAC9 protel
35	623.5	11.7	673	23 AAE22854	Human histone deac
36	623.5	11.7	673	21 AAB42743	Human ORFX ORF2507
37	623.5	11.7	686	21 AAB42757	Human ORFX ORF2521
38	580.5	10.9	330	22 AAB23410	Novel human enzyme
39	574.5	10.7	625	23 AAB81615	Human HDAC9 vari
40	527	9.9	363	22 AAO00244	Human histone deac
41	509.5	9.5	184	22 AAB94652	Human protein sequ
42	465	8.7	164	21 AAB43314	Human ORFX ORF3078
43	464.5	8.7	170	22 AAB95880	Human protein sequ
44	445	8.3	112	22 AAB49959	Human histone deac
45	374.5	7.0	335	20 AAM97110	Thermostable polyp

## ALIGNMENTS

RESULT 1  
AAB43008 standard; Protein: 1141 AA.  
AAB43008:  
08-FEB-2001 (first entry)

Human: open reading frame; ORFX; detection; cytosolic; hepatotropic;  
vulnerable; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;  
anticonvulsant; osteopontin; antitumor; immunosuppressant; cardiant;  
immunostimulant; thrombolytic; coagulant; vasodilator; antidiabetic;  
hypotensive; dermatological; immunosuppressive; antiinflammatory;  
antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;  
antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
neurodegenerative disorder; osteoarthritis; graft vs host disease;  
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;

KM cholesterol ester storage; systemic lupus erythematosus; infection;  
KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
KM bone damage; cartilage damage; antiinflammatory disease; coagulation;  
KM thrombosis; contraceptive.  
XX  
OS Homo sapiens.  
PN MO200058473-A2.  
XX  
PD 05-OCT-2000.  
XX  
PF 31-MAR-2000, 2000MO-US08621.  
XX  
PR 31-MAR-1999; 99US-0127607.  
XX  
PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX  
PA (CURA-) CURAGEN CORP.  
XX  
PI Shinkets RA, Leach M;  
XX  
DR WPI: 2000-602362/57.  
DR N-PSDB; AAC7217.  
XX  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX  
XX Claim 11: Page 4728-4731; 5507PP; English.  
XX  
XX AAC7446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulneryary;  
CC antiproliferative; antiparkinsonian; nootropic; neuroprotective;  
CC osteoplastic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; antineoplastic;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antihemetic;  
CC antihypertensive; and antianemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins. In gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX  
SQ Sequence 1141 AA:  
  
Alignment Scores:  
Pred. No.: 3,82e-315 Length: 1141  
Score: 4798.00 Matches: 927  
Percent Similarity: 99.57% Conservative: 3  
Best Local Similarity: 99.25% Mismatches: 4  
Query Match: 89.75% Indels: 1  
DB: 21 Gaps: 0  
  
US-09-502-945-2 (1-2885) x AAB43008 (1-1141)  
QY 2 GAATTCCTTGTGCAAGTCAAGAGCCACACCGGCGCTCAACCATTCCTCCCA 61  
DB 208 GlnPheLeuLeuSerLysSerLysGluProThrProGlyGlyLeuAsnHisSerLeuPro 227  
QY 62 CAGCACCCCAAAATGCTGGGAGCCACCATGCTTTGGACGAGTTCCCTCCCGAG 121  
DB 228 GlnHisProLysCysTrpGlyAlaHisHisAlaSerLeuAspGlnSerProProGln 247

QY 122 AGCGGCCCTCCCTGGAGCCCTCCCTCTACAACTGCTTTGGCTGGCCCTACGACAGT 181  
DB 248 SerGlyProProGlyLyrProProSerTyrLysLeuProLeuProGlyProTyrAspSer 267  
QY 182 CGAGACGACTTCCCTCCGCAAAACAGCCTCTGAACCCACTTTGAATGGCTTCAAG 241  
DB 268 ArgAspAspPheProLeuAlaGlyThrAlaSerGlnProAsnLeuLysValArgSerArg 287  
QY 242 CTAAGACAGAGAGGTGGCTGGAGAGAGAGCCCTCCCTGGCTGGCAAGGTGGAGCT 301  
DB 288 LeuLysGlnLysValAlaGluArgArgSerProLeuLeuArgArgLysAspGlyThr 307  
QY 302 GTTATTAGACCTTTAAGAAGAGAGCTGTGAGATCAAGGTGCGCGCTGGGGCTGCG 361  
DB 308 ValIleSerThrPheLysLysArgAlaValGluIleThrGlyAlaGlyProGlyAlaSer 327  
QY 362 TCCGTGTGTACAGCGCACCGGCTCCGGCCCGCAGCTTCCCAACAGCTCCACAGACC 421  
DB 328 SerValLysAsnSerAlaProGlySerGlyProSerSerProAsnSerHisSerThr 347  
QY 422 ATGCGTGAAGATGGCTTTACTGCTCAGTCCAGTCCCAATGCCCTGAGATGCTCCCTCG 481  
DB 348 IleAlaGlnAsnGlyPheThrGlySerValProAsnIleProThrGlnLeuLeuProGln 367  
QY 482 CACGAGCCCTCCCTCGGACAGCTCCCCCAACAGCTTGAAGCTTCAAGCTTCCCTCT 541  
DB 368 HisArgAlaLeuProLeuAspSerSerProAsnGlnPheSerLeuThrThrSerProSer 387  
QY 542 CTGCCCAACATCTCCCTAGGCTGCGAGCCGCGTCACTGTCCACCACTCACTCACT 601  
DB 388 LeuProAsnIleSerLeuGlyLeuGlnAlaThrValThrValThrAsnSerHisLeuThr 407  
QY 602 GCTTCCCGAAGGTGTGCGACACAGAGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 661  
DB 408 AlaSerProLysLeuSerThrGlnGlnGlnAlaGluArgGlnAlaLeuGlnSerLeuArg 427  
QY 662 CAGGTGGCAGCGTGAACGCGCAAGTTCATGAGACATCTCTATTCCTGAGCTGCTG 721  
DB 428 GlnGlyGlyThrLeuThrGlyLysPheMetSerThrSerSerIleProGlyCysLeuLeu 447  
QY 722 GCGGTGGCACTGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 781  
DB 448 GlyValAlaLeuGlnLysPheGlySerProHisGlyHisAlaSerLeuLeuGlnHisVal 467  
QY 782 CTTGCTCTGGAGAGGCGCGGAGAGACACCTCATTCCTGCTGCTGCTGCTGCTGCTG 841  
DB 468 LeuLeuLeuGlnLysAlaArgGlnGlnSerThrLeuIleAlaValProLeuHisGlyGln 487  
QY 842 TCCGCACATGATGACGGGTGAACGTGTGGCCACACGATGCGGAGGAGGAGGAGGAGG 901  
DB 488 SerProLeuValThrGlyLysAlaArgValAlaThrSerMetArgThrValGlyLysLeuPro 507  
QY 902 CGGCATCGGCCCTTGAGCCGACATCACTGCTCACCCTGCGGAGAGTCCCAAGGCCCTG 961  
DB 508 ArgHisArgProLeuSerArgThrGlnSerSerProLeuProGlnSerProGlnAlaLeu 527  
QY 962 CAGCAGCTGTCATGCAACAAACAGCAGCAGCAGCTCTGAGAGAGAGAGAGAGAGAGAG 1021  
DB 528 GlnGlnLeuValMetGlnGlnGlnHisGlnGlnPheLeuGlnLysGlnHisGlnGln 547  
QY 1022 CTACAGCTGGGCAAGATCTCTACCAAGAGAGGAGGAGCTCCAGGAGCCACACCCAG 1081  
DB 548 LeuGlnLeuGlyLysIleLeuThrLysThrGlnGlyLeuProArgGlnProThrThrHis 567  
QY 1082 CCTGAGAGACAGAGAGAGAGAGCTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1141  
DB 568 ProGlnGlnThrGlnGlnGlnLeuThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 587  
QY 1142 CTGACCATGCCCCGAGGAGGCTCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1201  
DB 588 LeuThrMetProArgGlnGlnGlnSerThrGlnSerGlnSerThrGlnGlnGlnGlnGln 607

[illegible]

Db	968	AlaHisGlnPheSerProAspValValLeuValSerAlaGlnPheAspAlaValGlnGly	967
Qy	2342	CAITGTCTCTCTCTGGGTGGCTACTCTGTGCACCGCCAGATGTTTGGCCACTTGACCAG	2401
Db	988	HisLeuSerProLeuGlyGlyTyrSerValThrAlaArgCysPheGlnHisLeuThrArg	1007
Qy	2402	CAGTGTATGACCTTGGCGAGGGGGCGGGTGTGTGGCCCTGGAGGGAGGCGATGACTTG	2461
Db	1008	GlnLeuMetThrLeuAlaGlnGlyArgValValLeuAlaLeuGlnGlnGlyGlnHisAspLeu	1027
Qy	2462	ACGGCCATCTGTATGCCCTCTGAAGTTTGTCTCTGGGCTCTGCTCAGTGTAAAGTGCAG	2521
Db	1028	ThrAlaIleCysAspAlaSerGlnAlaCysValSerAlaLeuLeuSerValGlnLeuGln	1047
Qy	2522	CCCTTGGATGAGCAGCTCTTGGCCAGCAAAGCCCAACATCAACGCGAGTGGCCAGCTAGAG	2581
Db	1048	ProLeuAspGlnAlaValAlaLeuGlnGlnGlnProAsnIleAsnAlaValAlaThrLeuGln	1067
Qy	2582	AAAGTCATCGAGATCCAGACCAACACTGTGAGCTGTGTGCAGAAATTGGCGGTGTCTG	2641
Db	1068	LysValIleGlnIleGlnSerIleThrSerCysValGlnLysPheAlaIleGlnLeu	1087
Qy	2642	GGCCGGTCCCTGGAGGGGCCCAAGAGGTGAGACCGAAGAGCCGAAT-GTGAACGCC	2700
Db	1088	GlyArgSerLeuArgGlnAlaGlnAlaGlnGlyGlnThrGlnAlaGlnIleThrValSerAla	1107
Qy	2701	ATGGCTCTTGGCTTGGTGGGGGGCCGAACAGGCCCAAGCTGGCGAGCCCGGGAACACAGC	2760
Db	1108	MetAlaLeuLeuSerValGlyAlaGlnGlnAlaGlnAlaIleAlaIleAlaIleArgGlnHisSer	1127
Qy	2761	CCCAAGCGCGGAGAGGAGCCCATGAGCAGAGGAGCCCTGGCCCTG	2802
Db	1128	ProArgProAlaGlnGlnProMetGlnGlnGlnProAlaIleu	1141
RESULT 2			
AA07092	AA07092 standard; Protein; 897 AA.		
AA07092:			
02-JUL-1999	(first entry)		
Colon cancer associated antigen precursor sequence.			
Cancer associated antigen: diagnosis; research; treatment; human;			
breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;			
prostate cancer.			
Homo sapiens.			
MO9904265-A2.			
28-JAN-1999.			
15-JUL-1998;	98WO-US14679.		
22-JUN-1998;	98US-0102322.		
17-JUL-1997;	97US-0896164.		
10-OCT-1997;	97US-0061599.		
10-OCT-1997;	97US-0061765.		
10-OCT-1997;	97US-0948705.		
11-OCT-1997;	97GB-0021697.		
(LUDM-) LUDMIG INST CANCER RES.			
Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;			
Pfundschn M, Sahin U, Scanlan MJ, Stockert E;			
Tureci O.			
WPI; 1999-132448/11.			
New isolated cancer associated nucleic acids and polypeptides -			



PT Isolated using sera from cancer patients, used to develop products  
PI for the diagnosis, monitoring or treatment of cancers  
XX  
PS Disclosure: Page 652-654; 787pp; English.

XX The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.

SQ Sequence 897 AA;

Alignment Scores:  
Pred. No.: 1.66e-306 Length: 897  
Score: 4669.00 Matches: 896  
Percent Similarity: 99.89% Conservative: 0  
Best Local Similarity: 99.89% Mismatches: 1  
Query Match: 87.34% Indels: 0  
DB: Gaps: 0

US-09-502-945-2 (1-2885) x AAY07092 (1-897)

OY 2 GAATTCCTTTGTCAGAAAGTCAAGAGAGCCACACAGGCGGCTCAACCTTCCCTCCA 61  
DB 1 GIuphLeuLeuSerLysSerLysGluProThrProGlyGlyLeuAsnHisSerLeuPro 20  
OY 62 CAGCACCCCAATGCTGGGAGCCACCACTGCTTTGGACAGATTCCCTCCACAG 121  
DB 21 GlnHisProLysScyTrpLysAlaHisHisAlaSerLeuAsnGlnSerSerProGln 40  
OY 122 AGCGGCGCCCTGGGAGCGCTCCCTCTCAAAAGCTGTCCTGGGCGCCACAGCACT 181  
DB 41 SerGlyProProGlyThrProProSerLysLysLeuProLeuProGlyPro\*\*AspSer 60  
OY 182 CGAGACAGACTCCCTCCGCAAAACAGCCTCTGAACCCCACTTGAATGGCTTCAAG 241  
DB 61 ArgAspAspPheProLeuArgLysThrAlaSerGluProAsnLeuLysValArgSerArg 80  
OY 242 CTAAACAGAGAGGTGGCTGAGCGAGAGACAGTCCCTCTCTGCGCAGAGATGGACT 301  
DB 81 LeuLysGlnLysValAlaGluArgArgSerSerProLeuLeuArgArgLysAspGlyThr 100  
OY 302 GTTATTAGACCTTTAAGAAGAGAGCTGTGATCATCAGGCGCGGCTGGGGGCTG 361  
DB 101 ValIleSerThrPheLysLysArgAlaValGluIleThrGlyAlaGlyProGlyAlaSer 120  
OY 362 TCCGCTGTGAACAGGAGCACCCTCCGCGCCAGCTCTCCCAAGAGCTCCACAGCACC 421  
DB 121 SerValLysAsnSerAlaProGlySerGlyProSerSerProAsnSerSerHisSerThr 140  
OY 422 ATCGCTGAGATGGCTTTACTGCTCAGTCCCAACATCCCACTGAGATGCTCCCTAG 481  
DB 141 IleAlaGluAsnGlyPheThrGlySerValProAsnIleProThrGluMetLeuProGln 160  
OY 482 CACGAGCGCCCTCCCTGGAGAGCTCCCGCAACAGTTAGCTTACAGCTCCTCTCT 541  
DB 161 HisArgAlaLeuProLeuAspSerSerProAsnGlnPheSerLeuLysThrSerProSer 180  
OY 542 CTGGCCCAATCTCCCTAGGGCTGAGGCGCAGCTCACTGTCACCAACTCAGACCTCACT 601  
DB 181 LeuProAsnIleSerLeuGlnLysLeuGlnAlaThrValThrValHisSerHisLeuThr 200  
OY 602 GCTCTCCCGAAGCTGTGCACACAGCAGAGGCGGAGAGGCGCTCCAGTCCCTGCGG 661

DB 201 AlaSerProLysLeuSerThrGlnGlnGlnAlaArgGlnAlaLeuGlnSerLeuArg 220  
OY 662 CAGGCTGGACCTTCACCGGCAAGTTCAATGACACATCCTCATTCGCGCTGCGG 721  
DB 221 GlnGlyGlyThrLeuThrThrLysPheMetSerThrSerSerLeuProLysLeuLeu 240  
OY 722 GCGGTGGCACTGAGAGGCGAGCGAGCCCAAGGCGCATGCTCCCTGCGCAGCATGTG 781  
DB 241 GlyValAlaLeuGlnGlyLysPoliSerProHisGlyHisAlaSerLeuLeuHisVal 260  
OY 782 CTGTTGCTGAGACAGGCGCGGACAGACACCTTCATTCCTGTCACCTCCAGCGCAG 841  
DB 261 LeuLeuLeuGlnGlnAlaAlaArgGlnGlnSerThrLeuIleAlaValProLeuHisGlyGln 280  
OY 842 TCCCCACTAGTAGAGGGGTGAAGCTGGCCACACAGCTCGGAGCGTGAAGCTCCCG 901  
DB 281 SerProLeuValThrGlyLysLysValAlaThrSerMetArgThrValLysLysLeuPro 300  
OY 902 CGGCAATCGGCCCTGAGCGGCACATGCTCCACCGCTGCGCCAGAGTCCCGGCGCTG 961  
DB 301 ArgHisArgProLeuSerThrArgThrGlnSerSerProLeuProGlnSerProGlnAlaLeu 320  
OY 962 CAGCAGCTGCTCATGCAACACAGCAGCAGATTCCTGAGAGAGACAGACAGCAG 1021  
DB 321 GlnGlnLeuValMetGlnGlnGlnHisGlnGlnPheLeuGlnLysGlnGlnGln 340  
OY 1022 CTACAGCTGGGCAAGATCCTCAACAAAGCAGGGAGCTGCCAGGACGCCACGCCAC 1081  
DB 341 LeuGlnLeuGlyLysLysLeuThrLysThrGlyGlyLeuLeuProArgGlnProThrThrHis 360  
OY 1082 CCTGAGAGACAGAGAGAGAGAGCTGACGAGCAGCAGAGAGTTCCTGGGGAGGAGGCC 1141  
DB 361 ProGlnGluThrGluGluGluLeuThrGluGlnGlnGlnValLeuLeuGlyLysAla 380  
OY 1142 CTGACCATGCGCCCGGAGGGGCTCCACAGAGATGAGACAGCAGACAGAACTGGAGAG 1201  
DB 381 LeuThrMetProArgGluLysLysSerThrGlnSerGlnSerThrGlnGlnLysPheGluGlu 400  
OY 1202 GAGCAGAGAGAGAGATGGGAGAGAGAGAGAGATTGCATCCAGTTAAGACAGCAGAG 1261  
DB 401 GluAspGluGluGluLysPoliLysLysGlnGluGluGluLysAspCysLysGlnValLysAspGluGlu 420  
OY 1262 GCGGAGAGTGTGCTGAGAGGGGCGCGACTTGAAGAGACCTGCTGCGATCAAAAAA 1321  
DB 421 GlyGlnSerGlyAlaGluGluGluLysProAspLeuGluGluProGlyAlaGlyLysLys 440  
OY 1322 CTGTTCTCAGATGCCCAACCGCTGCAACCTTGCAGGTGTACAGCGCCCTCAGCGCTG 1381  
DB 441 LeuPheSerAspAlaGlnProLeuGlnProLeuGlnValLysGlnAlaProLeuSerLeu 460  
OY 1382 GCCACTGTCCCGACCAAGCCCTGGGCGGTACCAATCCTCCCTGCTGCTCCCTGGGCGC 1441  
DB 461 AlaThrValProHisGlnAlaLeuGluLysArgThrGlnSerSerProAlaAlaProGlyGly 480  
OY 1442 ATGAAGAACCCCGACAGCAACCGCTCAAGCACTTCTCACCACAAAGTGTGTTACGAC 1501  
DB 481 MetLysAsnProProAspGlnProValLysHisLeuPheThrThrSerValValLysAsp 500  
OY 1502 ACGTTCAATGCTTAACACACAGTGCATGTGCGGGAACACACACAGCTGACCTGAGATGCT 1561  
DB 501 ThrPheMetLeuLysHisGlnLysMetCysGlyAsnThrHisValHisProGlnHisAla 520  
OY 1562 GCGCGATTCAGAGATGCTGTCGCGCTGCGAGAGACAGGCGCTGTTAGCAAGTGCAG 1621  
DB 521 GlyArgIleGlnSerIleThrSerArgLeuGlnGlnThrGlyLeuLeuSerLysCysGlu 540  
OY 1622 CGGATCCAGAGGTGCAAAAGCCAGCTAGATGATTCACAGACAGTCACTTGAATACAC 1681  
DB 541 ArgIleArgGlyArgGlyAlaThrLeuAspGluIleGlnThrValHisSerLysLysThrHis 560  
OY 1682 ACCTGTCTATAGGACCAAGTCCCTCAACCGGCAAGAGCTTAGACAGCAAGAGTTGCTC 1741

```

Db 561 ThrLeuLeuTyrglyThrSerProLeuAsnArgInLysLeuAspSerLysLeuLeu 580
QY 1742 GGTCCATCACCAGAGATGATGCTGTGCTGCTGTGGGGCATGGGGTGACAGT 1801
Db 581 GLYProLeuSerGlnLysMetTryAlaValLeuProCysGlyLysIleGlyValAspSer 600
QY 1802 GACACCGGTGGAATGAGATGCACACTCCTCCAGTGTGTGCGCATGGCAGTGGCTGTG 1861
Db 601 AspThrValTrpAsnGlnMetHisSerSerSerAlaValArgMetAlaValGlyCysLeu 620
QY 1862 CTGGAGCTGGGCTTCAGAGTGGCTGCAGAGAGACTCAAGATGGATTGGCATATCCGG 1921
Db 621 LeuGlnLeuAlaPheLysValAlaAlaGlyLysLeuLysAsnGlyPheAlaIleLeuArg 640
QY 1922 CCCCAGACACACAGCGGAGAGATCCACAGCATGGAGATTCTGCTTCACTGCTTA 1981
Db 641 ProProGlnHisHisAlaGlnGlnSerThrAlaMetGlyPheCysPhePheAsnSerVal 660
QY 1982 GCCATCACCGCAAAATCTCTACAGCAGAAGTTGAACGTGGGCAAGTCTCATGTGAC 2041
Db 661 AlaIleThrAlaLysLeuLeuGlnGlnLysLeuAsnValGlyLysValLeuIleValAsp 680
QY 2042 TGGGACATTCACCATGGCAATGSCACCCAGCGCTTTCACATGACCCCTGTGTCTC 2101
Db 681 TrpAspIleHisHisGlyAsnGlyThrGlnGlnAlaPheTyrrAsnAspProSerValLeu 700
QY 2102 TACATCTCTGCAATCGGTATGACAGCGGAATCTTCCAGGCTGTGGGGCTCCCTGAA 2161
Db 701 TyrlleSerLeuHisAlaGlyTrpAspAsnGlyAsnPhePheProGlySerGlyAlaProGlu 720
QY 2162 GAGGTGGTGGAGGACAGCGGTGGGTACATGTGAACGTGGCATGGATGGACAGAGTGTG 2221
Db 721 GlnValGlyGlyGlyProGlyValGlyTyrrAsnValAsnValAlaIleTrpHisGlyVal 740
QY 2222 GACCCCCCATTTGAGACAGTGGACTTACCTTACACCTTCAGAGACAGTGGTATGCCAT 2281
Db 741 AspProProlIleGlyAspValGlyLysLeuThrAlaPheArgThrValIleMetProIle 760
QY 2282 GCCCAGAGTCTGACCTGATGGTGTCTTACTGCTCCCGGGGTGATGCTGTGAAGGA 2341
Db 761 AlaHisGlnPheSerProAspValValLeuValSerAlaGlyPheAspAlaValGlnGly 780
QY 2342 CATCTGTCTCTCTGGGTGGTGTACTGTCTACCGCCAGATGTTTGGCCACTTGACAGG 2401
Db 781 HisLeuSerProLeuGlyGlyTyrrSerValThrAlaArgCysPheGlnHisLeuThrArg 800
QY 2402 CAGCTGATGACCTTGCGACGGGGCCGGGTGGTGTGGCTGGCTGGAAGGACCATGACTG 2461
Db 801 GlnLeuMetThrLeuAlaGlnGlyAlaGlyValValLeuAlaLeuGlnGlyLysHisAspLeu 820
QY 2462 ACCGCGCATGCTGATGGCTGTGAAGCTTGTGCTCCGGCTGCTGACGTAAAGCTGCAG 2521
Db 821 ThrAlaIleCysAspAlaSerGlnAlaCysValSerAlaLeuLeuSerValLysLeuGln 840
QY 2522 CCGTTGGATGAGCGAGTCTTGACAGCAAAAGCCCAACATCAAGCAGTGGCCACCTGAG 2581
Db 841 ProLeuAspGlnAlaValLeuGlnGlnLysProAsnIleAsnAlaValAlaThrLeuGln 860
QY 2582 AAAGTCATCGAGATCCAGAGCAAACTGAGCTGTGTGCGAAGTTGCCCGCTGTGTG 2641
Db 861 LysValIleGlnIleGlnIleGlnSerLysHisTrpSerCysValGlnLysPheAlaAlaGlyLeu 880
QY 2642 GCGCGGTCCTCGGAGGGGCCAGAGGTGAGACCGGAAGCGGGAATG 2692
Db 881 GlyArgSerLeuArgGlyAlaGlnAlaGlyLysIleThrGlnGlnAlaGlnMet 897

```

```

RESULT 3
AAB49958 standard; Protein; 716 AA.
ID AAB49958
AC AAB49958;
XX
XX 08-MAR-2001 (first entry)

```

```

XX XX Human histone deacetylase HDAC-5.
DE
XX XX
KW Histone deacetylase; HDAC-1; HDAC-2; HDAC-3; HDAC-4; HDAC-5; HDAC-C;
KW HDAC-D; cell cycle; tumorigenesis; cancer; inhibitor; antisense;
KW gene therapy.
XX
OS Homo sapiens.
XX
PN M0200071703-A2.
XX
PD 30-NOV-2000.
XX
PF 03-MAY-2000; 2000MO-IB01252.
XX
PR 03-MAY-1999; 99US-0132287.
XX
PA (METH-) METHYLENE INC.
PI Macleod AR, Li Z, Beerman JM;
PI WPI: 2001-016407/02.
DR N-PSDB; AAC89558.
DR
PT Antisense oligonucleotide that inhibits expression of a histone
PT deacetylase, useful for treating and/or alleviating the symptoms of
PT neoplasia, or for inhibiting neoplastic cell growth in an animal -
XX
PS Disclosure; Page 62-64; 125pp; English.
XX
CC The present invention provides inhibitors of histone deacetylase enzymes
CC such as HDAC-1, HDAC-2, HDAC-3, HDAC-4, HDAC-5, HDAC-C and HDAC-D. These
CC inhibitors may be antisense strands or they may be compounds identified
CC by contacting the enzyme with the compound and measuring the resulting
CC enzyme activity. These inhibitors are useful for treating cancers and for
CC identifying which histone deacetylase is involved in a neoplasia.
XX
SQ Sequence 716 AA;

```

```

Alignment Scores: 2.03e-238 Length: 716
Pred. No.: 3659.00 Matches: 709
Score: 99.44% Conservative: 3
Percent Similarity: 99.02% Mismatches: 4
Best Local Similarity: 68.44% Indels: 1
Query Match: 22 Gaps: 0

```

US-09-502-945-2 (1-2885) x AAB49958 (1-716)

```

QY 656 CTGGCGGAGGTGGTGGCAGCTGACCGGCAAGTTCATGAGCACATCTGATTCGGGTGC 715
Db 1 LeuArgGlnGlyGlyThrLeuThrGlyLysPheMetSerThrSerIleProGlyCys 20
QY 716 CTGCTGGCGGTGACCTGAGGGGAGCGAGGAGCCCGCCAGCGGCAATGCTCTGCTGCAG 775
Db 21 LeuLeuGlyAlaAlaLeuGlnGlyAspGlySerProHisGlnHisAlaSerLeuLeuGln 40
QY 776 CATGTGCTGTGCTGAGACAGGCGCGGCGAGAGCAAGCCATGCTGCTGTGCGCTCCAC 835
Db 41 HisValLeuLeuLeuGlnGlnAlaArgGlnIleSerThrLeuIleAlaValProLeuHis 60
QY 836 GGGCAGTCCCGCATGATGACGGGTGAACGTGTGCGCCACAGCATGGGAGGTGAGCAAG 895
Db 61 GlyIleSerProLeuValThrGlyGlnArgValAlaThrSerMetAlaGlnValGlyLys 80
QY 896 CTCCCGGCGCATGCGCCCTGAGCGGCACTGCTCCACCGGTGCCAGAGATCCCGCAG 955
Db 81 LeuProArgHisArgProLeuSerArgThrGlnSerSerProLeuProGlnInsProGln 100
QY 956 GCCCTGCAGCAGCTGTGTCATGCAACAACAGCAGCAGCAGTTCCTGGGAAGCAGAACAG 1015
Db 101 AlaLeuGlnGlnLeuValMetGlnGlnGlnHisGlnGlnPheLeuGlnLysGlnLysGln 120

```

QY 1016 CAGCAGCTACAGCTGGGCAAGATCTCACACAGAGGGGAGCTGCCACGAGCCACC 1075  
 |||||  
 Db 121 GlnGlnLeuGlnLeuGlnLysIleLeuThrLysThrGlyIuLeuProArgGlnProThr 140  
 QY 1076 ACCGACCTGAGAGACAGAGAGAGAGCTGACGAGACAGAGAGCTTGGTGGGGAG 1135  
 |||||  
 Db 141 ThrHisProGlnIuThrGlnGlnGlnLeuThrGlnGlnGlnIuValLeuGlnGlyIu 160  
 QY 1136 GAGACCCCTGACCAATGCCCGGAGGAGCTCCACAGAGATGAGAGCACAGCAAGACTG 1195  
 |||||  
 Db 161 GlyAlaLeuThrMetProArgGlnGlnGlnLysSerThrGlnSerIuSerThrGlnGlnAlaPhe 180  
 QY 1196 GAGGAGAGAGAGAGAGAGAGATGGGAGAGAGAGAGAGATTCATCCAGCTTAAGAC 1255  
 |||||  
 Db 181 GlnGlnGlnIuAspGlnGlnGlnIuAspGlnGlnGlnIuAspGlnGlnIuAspGlnGlnIu 200  
 QY 1256 GAGGAGGAGAGAGATGGTGGTGGAGAGGGGCGGACTGGAGAGAGCTGGTGGGTAC 1315  
 |||||  
 Db 201 GlnGlnGlnIuGlnSerGlnIuGlnGlnGlnGlnIuProAspLeuGlnIuProGlnIuAlaGly 220  
 QY 1316 AAAAAGCTGTCTCAGATGCCCAACCGCTGCACACTTTGAGGTGTACCAAGCCGCTC 1375  
 |||||  
 Db 221 LysLysLeuPheSerAspAlaGlnProLeuGlnProLeuGlnIuValIuGlnAlaProLeu 240  
 QY 1376 AGCTGGCCACTGTGCCCAACAGCCCTGGGCGGTACCAATCCCTCCCTGCTGCCCT 1435  
 |||||  
 Db 241 SerLeuAlaThrValIuProHisGlnAlaLeuGlnIuArgThrGlnSerProAlaAlaPro 260  
 QY 1436 GGGGAGATGAAGAACCCCGACAGACCAACCGCTGCACACTTTGAGGTGTACCAAGCTGTGTC 1495  
 |||||  
 Db 261 GlyGlnMetLysSerProProAspGlnProValIuHisLeuPheThrThrGlyValIu 280  
 QY 1496 TACGACACCTTCATGCTAAAGACACAGTGCATGGCGGGAACACACAGCTGCACCTGAG 1555  
 |||||  
 Db 281 TyrAspThrPheMetLeuLysHisGlnCysMetCysGlnIuAsnThrHisValHisProGln 300  
 QY 1556 CATTGGCGCGGATCCAGAGCAATCTGCTCCCGGCTGCAGAGACAGCCCTGTACCAAG 1615  
 |||||  
 Db 301 HisAlaGlnIuArgIleGlnSerIleIuTrpSerArgLeuGlnIuThrIuLeuLeuSerLys 320  
 QY 1616 TGGCAGCGATCCGAGGTGCGCAAGCCAGCTGATGAGATCCAGACAGTCACTGTGA 1675  
 |||||  
 Db 321 CysGlnIuArgIleArgIuArgLysAlaThrLeuAspGlnIleGlnThrValHisSerGln 340  
 QY 1676 TACACACCCCTGCTATGAGGAGCAAGTCCCTGCACCCGAGAGAGTGAACAGCAAGAG 1735  
 |||||  
 Db 341 TyrHisThrLeuLeuLysArgIuThrSerProLeuAsnArgIuGlnLysLeuAspSerLys 360  
 QY 1736 TTGCTGGGTCCCATCAGCCAGAAAGATGTATGCTGCTGCTGCTGGGGCATCGGGGTG 1795  
 |||||  
 Db 361 LeuLeuGlnIuProIleSerGlnLysMetIuValIuValLeuProCysGlnIuGlnIuVal 380  
 QY 1796 GACAGTGACACCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1855  
 |||||  
 Db 381 AspSerAspThrValIuTrpAsnGlnMetHisSerSerSerAlaValIuArgMetAlaValIuGly 400  
 QY 1856 TTGCTGGGTCCCATCAGCCAGAAAGATGTATGCTGCTGCTGCTGGGGCATCGGGGTG 1915  
 |||||  
 Db 401 CysLeuLeuGlnIuLeuAlaPheLysValAlaIaGlnIuLeuLysAsnGlnIuPheAlaIle 420  
 QY 1916 ATCCGGCCCCCAGAGACACAGCCGAGAGATGCACAGCCATGGGATTCGCTTCTTCAAC 1975  
 |||||  
 Db 421 IleArgProProGlnIuHisHisAlaGlnIuSerThrAlaMetIuLysPheCysPhePheAsn 440  
 QY 1976 TCTGTAGCATCAGCGCAAAACGCTACAGCAAGTGAAGTGAAGTGGGCAAGGCTCTCAGC 2035  
 |||||  
 Db 441 SerValAlaIleThrAlaLysLeuLeuGlnGlnLysLeuAsnValGlnLysValLeuIle 460  
 QY 2036 GTGAGCTGGAGATTACCATGGGCAATGGCAGCCAGCCAGGCTTACATGACCCCTCT 2095  
 |||||  
 Db 461 ValAspTrpAspIleHisHisGlnGlnGlnGlnIuThrGlnGlnAlaPheLysTrpAsnProSer 480  
 QY 2096 GTGCTCTACATCTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2155  
 |||||

Db 481 ValLeuTyrIleSerLeuHisArgTyrAspAsnGlnIuAsnPheProGlnSerGlyAla 500  
 QY 2156 CCGTAGAGGCTGGTGGAGAGCAGGCGGGGTACAAATGTGAACGCGCATGAGAGAG 2215  
 |||||  
 Db 501 ProGlnIuValIuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520  
 QY 2216 GGTGTGAGCCCCCATTTGAGAGAGTGAAGTACCTTACAGCCCTTACAGAGAGTGTATG 2275  
 |||||  
 Db 521 GlyValAspProProIleGlnAspValIuGlnIuLeuThrAlaPheArgThrValIuMet 540  
 QY 2276 CCGATTGCCACAGATTCTACCTGATGTGCTGTACCTGTCCCGCGGTTTATGCTGT 2335  
 |||||  
 Db 541 ProIleAlaHisGlnPheSerProAspValIuValLeuValSerIaGlnIuPheAspAlaVal 560  
 QY 2336 GAGGACATCTGCTCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2395  
 |||||  
 Db 561 GlnGlnIuHisLeuSerProLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 580  
 QY 2396 ACCAGGACGCTGATGACCTTGGCAGGGGGCGGGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2455  
 |||||  
 Db 581 ThrArgGlnLeuMetThrLeuAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 600  
 QY 2456 GACTTACCGCCATCTGTGTATGCTGTGAAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2515  
 |||||  
 Db 601 AspLeuThrAlaIleCysAspAlaSerGlnAlaCysValSerIaLeuLeuSerValGln 620  
 QY 2516 CTGCAGCCCTTGGATGAGCAGTCTTGCAGCAAAAGCCCAACATCAACGACAGTGCAGAG 2575  
 |||||  
 Db 621 LeuGlnProLeuAspGlnIuAlaValLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 640  
 QY 2576 CTAGAGAAAGTCAATGATCCAGAGCAAAAGTGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2635  
 |||||  
 Db 641 LeuGlnIuValIuIleGlnIuLeuGlnSerLysHisIuTrpSerCysValGlnIuLysPheAlaIa 660  
 QY 2636 GGTGTGAGCGCGCTTGGTGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2694  
 |||||  
 Db 661 GlyLeuGlnIuArgSerLeuArgIuAlaGlnAlaGlnIuGlnIuGlnIuGlnIuGlnIuGlnIu 680  
 QY 2695 AACGCGATGGCTGT 2754  
 |||||  
 Db 681 SerAlaMetAlaLeuLeuSerValIuGlnIuGlnIuAlaGlnIuAlaIaIaIaIaIaIaIaIaIaIa 700  
 QY 2755 CACAGCCCCAGGCGGCGAGAGAGCCCATGAGACAGAGAGCTGCCCTG 2802  
 |||||  
 Db 701 HisSerProArgProAlaGlnIuProMetGlnGlnIuProAlaLeu 716  
 RESULT 4  
 AA099660  
 ID AA099660 standard; Protein; 717 AA.  
 AC AA099660;  
 XX  
 DT 07-OCT-2002 (first entry)  
 DE Human Histone deacetylase isoform 5.  
 XX  
 KW Human; histone deacetylase; HDAC-5; cancer; cytostatic;  
 KW antisense; tumour suppressor; cell proliferation; tumour;  
 KW programmed cell death; necrotic cell death.  
 OS Homo sapiens.  
 OS  
 FH Key  
 FH Location/Qualifiers  
 FT Misc-difference 102..193  
 FT /note= "Encoded by GAGGAGAT"  
 FT Misc-difference 340..342  
 FT /note= "Encoded by TACCAC"  
 FT Misc-difference 434  
 FT /note= "Encoded by ACG"  
 FT Misc-difference 551..553  
 FT /note= "Encoded by GTCTTA"  
 FT Misc-difference 628

FT /note= "Encoded by GCA"  
FT Misc-difference 648  
FT /note= "Encoded by ATC"  
FT Misc-difference 675  
FT /note= "Encoded by GCC"  
XX  
PN US2002061860-A1.  
XX  
PD 23-MAY-2002.  
XX  
PF 06-AUG-2001; 2001US-0817913.  
XX  
PR 24-MAR-2000; 2000US-192157P.  
XX  
PA (LIZZ/) LI Z.  
PA (BONF/) BONFITIS C.  
PA (BEST/) BESTERMAN J.  
XX  
PI LI Z, Bonfils C, Besterman J;  
XX  
DR MPI: 2002-507650/54.  
DR N-PsDB; ABK87719.  
XX  
PT Agent that specifically inhibits an isoform of histone deacetylase,  
PT useful for treating cancer and other cell proliferative diseases,  
PT preferably comprises an antisense oligonucleotide -  
XX  
PS Disclosure; Fig 5A; 60pp; English.  
XX  
CC The invention relates to an agent that inhibits an isoform of histone  
CC deacetylase (HDAC-1 to HDAC-8) but not all isoforms, e.g. an  
CC antisense oligonucleotide. Also included are inhibiting an HDAC isoform  
CC in a cell by treatment with the agent, identifying an HDAC isoform that  
CC is required for induction of cell proliferation or differentiation and  
CC inhibiting cell proliferation by treatment with two antisense  
CC oligonucleotides or small molecules that inhibit a specific HDAC  
CC isoform, or antisense oligonucleotide or small molecules that inhibit  
CC DNA methyltransferase. The agent therefore acts as a tumour suppressor.  
CC The agents are used to treat diseases of cell proliferation and  
CC differentiation (e.g. cancer and tumours), by inducing growth retardation,  
CC growth arrest or programmed/necrotic cell death, specifically neoplastic  
CC cell proliferation in humans. The agents are selective for particular  
CC isoforms, compared to known inhibitors which are not selective.  
CC  
CC The present sequence represents the HDAC-5 isoform.  
XX  
SQ Sequence 717 AA:

Alignment Scores:

Pred. No.: 2, 77e-235 Length: 717  
Score: 3612.50 Matches: 706  
Percent Similarity: 98.75% Conservat: 3  
Best Local Similarity: 98.33% Mismatches: 6  
Query Match: 67.57% Indels: 4  
DB: 23 Gaps: 3

US-09-502-945-2 (1-2885) x AA099660 (1-717)

QY 656 CTCGCGAGGGGTGGACGGTGCAGACCGCAATTCATGAGCATCTCTATTCTGGTGC 715  
DB 1 LeuArgGln1yglThrLeuThrGlyLysPheMetSerThrSerSerIleProGlyCys 20  
QY 716 CTGCTGGGCTGGCAGTGGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 775  
DB 21 LeuLeuGlyAlaAlaLeuGlnGlyLysPglySerProHisIleGlnHisLeuLeuGln 40  
QY 776 CATTGCTGTGTGTGAGAGAGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 835  
DB 41 HisValLeuLeuLeuGlnGlnAlaArgGlnGlnSerThrLeuIleAlaValProLeuHis 60  
QY 836 GGGCAGTCCCTAGTGAAGAGGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAGTGAAGAG 895  
DB 61 G1yGlnSerProLeuValThrGlyGlnArgValAlaThrSerMetArgThrValGlyLys 80

QY 896 CTCGCGGAGTGGAGCCCTGAGCCGAGCATGATCTCTACCGGTGGCGAGAGTCCCGAG 955  
DB 81 LeuProArgHisThrProLeuSerArgThrGlnSerSerProLeuProGlnInsProGln 100  
QY 956 GCCCTGAGCAGAGCTGTGATGCAACAACAGACAGACAGAGTCTCTGGAGAAGAGAGAG 1015  
DB 101 AlaLeuGlnGlnLeuValMetGlnGlnGlnHisGlnGlnPheLeuGlnLysGlnGln 120  
QY 1016 CAGCAGCTACAGCTGGGCAAGATCTTCACCAACAGAGGAGAGTGGCCGAGGAGCCAGC 1075  
DB 121 GlnGlnLeuGlnLeuGlnLysIleLeuThrLysThrGlyLeuLeuProArgGlnProThr 140  
QY 1076 ACCCAGCTGAGGAG 1135  
DB 141 ThrHisProGlnGlnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 160  
QY 1136 GGAAGCCCTGAGCATGCCCCGGAGGGGCTCCAGAGAGTGAAGACACAGAGAGAGAGCTG 1195  
DB 161 G1yAlaLeuThrMetProArgGlnGlySerThrGlnSerGlnSerThrGlnGlnLysPleu 180  
QY 1196 GAG 1255  
DB 181 GlnGlnLeuLysPglyGlnGlnLysPgly--GlnGlnLeuLysPglyIleGlnValLysAsp 199  
QY 1256 GAG 1315  
DB 200 GlnGlnLysLysLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 219  
QY 1316 AAAAAGCTGTTCTCAGATGCCCAACCGCTGCAACCTTTCAGAGTGTACCAAGGCCCTTC 1375  
DB 220 LysLysLeuPheSerArgPalaGlnProLeuGlnProLeuGlnValTyrglnAlaProLeu 239  
QY 1376 AGCTGGCCACTGTGCCCCCAAGAGCCCTGGCCGTCATCCCAATCTCCCTGCTGCCCT 1435  
DB 240 SerLeuAlaThrValProHisGlnAlaLeuGlnValArgThrGlnSerProAlaAlaPro 259  
QY 1436 GGGGGCATGAAGAACCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1495  
DB 260 GlyGlyMetLysSerProPalaGlnProValLysHisLeuThrThrGlyValVal 279  
QY 1496 TACGACACGTTGATGCTAAAGACACAGTGCATGTGGGAGAACACACAGTGCACCTGAG 1555  
DB 280 TyrAspThrPheMetLeuLysHisGlnCysMetCysGlyAsnThrHisValHisProGln 299  
QY 1556 CATTGCTGGCGGATCCAGACAGATCTGGTCCGGCTGGAGAGAGAGAGAGAGAGAGAG 1615  
DB 300 HisAlaGlyValArgIleGlnSerIleTyrPserArgLeuGlnGlnThrGlyLeuLeuSerLys 319  
QY 1616 TGGCAGCGAGATCCGAGGTCGCAAGCCAGCTGATGATGATCCAGAGAGTGCACCTGAA 1675  
DB 320 CysGlnArgIleArgGlyArgGlyAlaThrLeuAspGlnIleGlnThrValHisSerGln 339  
QY 1676 TAC--CACACCTGCTCTATGAGACAGTCCCTCAACCGGAGAGAGAGTGAACAGCAG 1732  
DB 340 TyrIleHisThrLeuLeuTyrglyTyrSerProLeuAsnArgGlnLysLeuAspSerLys 359  
QY 1733 AAGTTGCTGGTCCCATCAACCCAGAGATGATGATGATGATGATGATGATGATGATGATG 1792  
DB 360 LysLeuLeuGlnProIleSerGlnLysMetTyraAlaAlaLeuProCysGlyGlyIleGly 379  
QY 1793 GTGAGCAGTGAACCGTGTGAATGAGATGATGATGATGATGATGATGATGATGATGATG 1852  
DB 380 ValAspSerAspThrValIleTyrAsnGlnMetHisSerSerSerAlaValArgMetAlaVal 399  
QY 1853 GGTGCTGCTGAGAGTGGCTTCAAGTGGTGCAGAGAGAGTCAAGAAATGATTTGCC 1912  
DB 400 GlyCysLeuLeuGlnLeuAlaPheLysValAlaAlaGlyLysLeuLysAsnGlyPheAla 419  
QY 1913 ATCATCGGCGCCGAGAGACACAGCGGAGAGATTCACAGAGAGAGAGAGAGAGAGAGAG 1972  
DB 420 IleIleArgProProGlnHisHisAlaGlnGlnGlnSerThrAlaMetCysGlyPheCysPhePhe 439  
QY 1973 AACCTGTAGCCATCAACCGCAAACTCTTACAGCAGAACTTGAACGTGGGAGAGTCTCTC 2032

```
DB 440 AasnSerValAlaIleThrAlaLysLeuGlnGlnLysLeuAsnValGlyLysValLeu 459
OY 2033 ATCGGAGCATGGACATTACCATGCAATGGACACCCAGCCGCTTACATGACCCC 2092
DB 460 ILevalAspTTPAspIleHisIshISglYasnGlyThrGlnGlnAlaPheThyrAsnAspPro 479
OY 2093 TCTGTGCTCTACATCTCTCTGCACTATGCAACAGGAGACTTCTTCCAGGCTCTGGG 2152
DB 480 SerValLeuTyrlIleSerLeuHisArgTyrAspAsnGlyAsnPheProGlySerGly 499
OY 2153 GCTCCTGAAGAGCTTGTTGGAGAGACAGCGCTGGGCTAACATGTGACCTGGCATGGACA 2212
DB 500 AlaProGlnGlnValGlyGlyProGlyValGlyTyrAsnValAsnValAlaTTPThr 519
OY 2213 GGAGGTGTGAGCCCCCCTTGGAGACGTGAGTACCTACAGCCCTTGAGACAGTGGTG 2272
DB 520 GlyGlyValAspProProIleGlyAspValGlyTyrLeuThrAlaPheArgThrValVal 539
OY 2273 ATGCCCATTTGCCACGAGTTCTACCTGATGTGTC--CTAGTCTCCGCGGGTTGAT 2329
DB 540 MetProIleAlaHisGlnPheSerProAspValValThrLeuValSerAlaGlyPheAsp 559
OY 2330 GCTGTGGAAGACATCTGCTCTCTGCTGGGCTGCTACTCTGTACCGCCGATGTTTGGC 2389
DB 560 AlaValAlGlnGlyHisLeuSerProLeuGlnGlyTyrSerValThrAlaArgCysPheGly 579
OY 2390 CACTTGACACGAGCAGCTGATGACCCCTGGCAGGGGCGGCTGCTGACCCCTGGAGGGA 2449
DB 580 HisLeuThrArgGlnLeuMetThrLeuAlaGlyGlyArgValValLeuAlaLeuGlnGly 599
OY 2450 GGCATGATGATGACCGCCATCTGTATGAGCTGTGAAGCTTGTCTGAGCTCTGCTCACT 2509
DB 600 GlyHisAspLeuThrAlaIleCysAspAlaSerGlnAlaCysValSerAlaLeuLeuSer 619
OY 2510 GTAAAGCTGCACCCCTTGATGATGAGCAGCTTGTGCAGCAAAACCCCAACTCAACGAGTG 2569
DB 620 ValGlnLeuGlnProLeuAspGlnLeuValLeuGlnGlnLysProAsnIleAsnAlaVal 639
OY 2570 GCCAGCGTGAAGAAAGTCTATCGAGATCCAGAGCAAAACACTGGAGCTGTGCAAGATGC 2629
DB 640 AlaThrLeuGlnLysValIleGlnThrGlnSerLysHisTyrSerCysValGlnLysPhe 659
OY 2630 GCCGCTGTGCTGGCGCGCTCCCTGCGAGGGGCCCAAGCAGGTGAGACCGCAAGAGCCGAA 2689
DB 660 AlaAlaGlyLysGlyArgSerLeuArgGlnAlaGlnAlaGlyGlnThrGlnGlnAlaGln 679
OY 2690 AT-GTGAACGCCATGGCTTGTGTTGGTGGGGCGCAACAGCCCAAGCTGCGGACGC 2748
DB 680 ThrValSerAlaMetAlaLeuLeuSerValGlyAlaGlnGlnAlaGlnAlaAlaAla 699
OY 2749 CGGGAACACAGCCCGAGCGCGAGAGAGCCATGAGACAGAGAGCTGCCCTG 2802
DB 700 ArgGlnHisSerProArgProAlaGlnGlnPrometGlnGlnPromAlaLeu 717

RESULT 5
AAB49957 standard: Protein: 967 AA.
AC AAB49957;
AT 08-MAR-2001 (first entry)
DE Human histone deacetylase HDAC-4.
DX
KW Histone deacetylase; HDAC-1; HDAC-2; HDAC-3; HDAC-4; HDAC-5; HDAC-C;
KW HDAC-D; cell cycle; tumorigenesis; cancer; inhibitor; antisense;
KW gene therapy.
OS Homo sapiens.
XX
XX WO200071703-A2.
XX
```

```
PD 30-NOV-2000.
XX
XX 03-MAY-2000; 2000WO-IB01252.
XX
XX 03-MAY-1999; 99US-0132287.
XX
XX (METH-) METHYLENE INC.
XX
XX Macleod AR, Li Z, Besterman JM;
XX
XX MPI: 2001-016407/02.
XX
XX N-PSDB: AAC89557.
XX
XX Antisense oligonucleotide that inhibits expression of a histone
XX deacetylase, useful for treating and/or alleviating the symptoms of
XX neoplasia, or for inhibiting neoplastic cell growth in an animal -
XX
XX Disclosure: Page 56-59; 125pp; English.
XX
XX
XX The present invention provides inhibitors of histone deacetylase enzymes
XX such as HDAC-1, HDAC-2, HDAC-3, HDAC-4, HDAC-5, HDAC-C and HDAC-D. These
XX inhibitors may be antisense strands or they may be compounds identified
XX by contacting the enzyme with the compound and measuring the resulting
XX enzyme activity. These inhibitors are useful for treating cancers and for
XX identifying which histone deacetylase is involved in a neoplasia.
XX
XX
XX Sequence 967 AA:
XX
XX Alignment Scores:
XX Pred. No.: 4,24e-173 Length: 967
XX Score: 2691.00 Matches: 568
XX Percent Similarity: 71.38% Conservative: 118
XX Best Local Similarity: 59.11% Mismatches: 195
XX Query Match: 50.34% Indels: 81
XX DB: 22 Gaps: 23
XX
XX US-09-502-945-2 (1-2885) x AAB49957 (1-967)
OY 2 GAATTCCTCTGTGCAAAAGTCAAAAGAGGCCACAGCGGCGCTCAACCATTCCTCCGCA 61
DB 60 GlnPheValLeuAsnLysLysLysAlaLeuAlaHisArgAsnLeuAsnHisCysIleSer 79
OY 62 CAGCACCCCAAAATGCTGG-----GGAGCCCAACATGCTTGTGGACCGAGATTCCCT 115
DB 80 SerAspProArgTyrTyrPyrGlyThrGlnHisSerSerLeuAspIleSerSerPro 99
OY 116 CCCCAGAGCGGCCCCCTCCGGAAGCGCTCCCTCTACAAACTGCTTGGCTGGCGGCTCAG 175
DB 100 ProGlnSer-----GlyValSerThrSerTyrAsnHisProAlaLeuGlyMetTyr 116
OY 176 GACAGTCGAGACGACTTCCCTCCGCAAAACAGCTCTGAACCCCAACTTGAAGTGCST 235
DB 117 AspAlaLysAspAspPheProLeuArgLysThrAlaSerGlnProAsnLeuLysLeuArg 136
OY 236 TCAAGGCTAAACAGAAAGTGCCTGAGCGAGAGACAGTCCCTCTGCTGGCAAGAT 295
DB 137 SerArgLeuLysGlnLysValAlaGlnArgArgSerSerProLeuLeuArgLysAsp 156
OY 296 GGGAGCTGTTATTAGCACTTTTAAAGAGAGAGCTGTGAGATGACAGGTCCGGGCTGGG 355
DB 157 GlyProValValThrAlaLeuLysLysArgProLeuAspValThr----- 171
OY 356 GCGTGTCCGCTGTAAACAGCGCACCGGCTCCGGGCCAGGCTTCCC--AACAGCTCC 412
DB 172 ---AspSerAlaCysSerSerThrAlaProGlySerGlyProSerSerProAsnHisSer 190
OY 413 CACAGCACCATGCTGAGATGAGCTTTTACTGGCTCAGTCCCAACATCCCACTGAGATG 472
DB 191 GlySerValSerAlaGlnAsnGlyIleAlaProAlaValProSerIleProAlaGlnThr 210
OY 473 CTCCTCAGCACCGAGCCCTCTGAGACACTCCCCCAACAGTCAAGCTTACAG 532
DB 211 SerLeuAlaHisArgLeuValAlaArgGlnGlySerAlaAlaProLeuProLeuTyThr 230
```

QY 533 TFCGTTTCGSSCAACATCTCCSTAGAGGCGTGGAGGCGACGGTCACTGTCCACCAACATCA 592  
 |||||  
 Db 231 SerProSerLeuProAlaHisLeuLeuLeuLeuProAlaThr----- 244  
 QY 593 CACCTCACTGCTCCCGGAGCTGTGACA-----CAGCAGAGAGCCGAGAGCGAGGCC 646  
 |||||  
 Db 245 -----GlyProSerAlaGlyThrAlaGlyInGlnAspThrGlnAlaLeuThr 260  
 QY 647 CTCGACGTCCCTGGCGGAG-----GGTGGCAGCTGACCGGCACTTCATG 691  
 |||||  
 Db 261 LeuProAlaLeuGlnGlnAlaArgLeuSerLeuProGlyThrGlnHisLeuThrProTyrLeu 280  
 QY 692 AGCAGATCTCATATTCTGCTGCTGCTGGCGGCGAGCGAGGCGGAGGCGGAGGCC 751  
 |||||  
 Db 281 SerThrSer-----ProLeuGlnAlaArgAspGlyAla 291  
 QY 752 CACGGGCAATGCTCCCTGCTGACAGATGCTGTGTGTGAGCAGAGGCCGCGGACAGAGAC 811  
 |||||  
 Db 292 ---AlaHisSerProLeuGlnHisMetValLeuLeuGlnProProAlaGlnAla 310  
 QY 812 ACCCTCAT-----GCTGTGCACTCCAGGGGCACTCCCACTAGTGACGGGT 859  
 |||||  
 Db 311 ProLeuValThrGlyLeuGlyAlaLeuProLeuHisAlaGlnSer---LeuValGlyAla 329  
 QY 860 GAAGGTGTGGCCAGCAGATGCGGACGGTAGGCAAGCTCCCGGAGATCGGCGCCCTGAGC 919  
 |||||  
 Db 330 AspArgValSerProSer-----IleHisTyrLeuAlaGlnGlnHisArgProLeuGly 346  
 QY 920 CGCAGCTCAGTCTTACCGCTGCGCGGAGAGTCCCGAGGCCCTGACAGCATGTGATGCAG 979  
 |||||  
 Db 347 ArgThrGlnSerAlaProLeuProGlnAsnAlaGlnAlaLeuGlnHisLeuValIleGln 366  
 QY 980 CAACAGACACCGAGCTTCTGAGAGACAGAG-----CAGCAGCAGCTATAG 1027  
 |||||  
 Db 367 GlnIleHisGlnGlnProLeuGlnHisTyrGlnGlnIlePheGlnGlnGlnIleGln 386  
 QY 1028 CTGGGCAAGATCTTACCAACAGAGAGGAGCTGCCAGGCGCCACCGACCATCTGAG 1087  
 |||||  
 Db 387 MetAsnTyrIleLeuProLysProSerGlnProAlaArgIleProGlnSerHisProGlu 406  
 QY 1088 GAGACAGAGAGAGAGCTGACGAGCAGCAGCAGAGAGCTTGTGCGGGAGAGAGCCCTG--- 1144  
 |||||  
 Db 407 GlnThrGlnGlnGlnIleuAlaArgGlnHisGln---AlaLeuAspArgGluProTyrLeuAsp 425  
 QY 1145 ASCATGCCCCGGAGGGGCTCCACAGAGAGTGAGAGACACAGAGAAAGCCTGAGAGAGAG 1204  
 |||||  
 Db 426 ArgLeuProGlyGlnLysGlnAlaHisAlaGlnAlaGlyAlaGln---ValLysGlnGlu 444  
 QY 1205 GACGAGAGAGAGATGGGAGAGAGAGAGATTCATCCAGGTTAAGAGCAGAGAGAGCGC 1264  
 |||||  
 Db 445 ProIleGlnSerAspGlnGlnAlaGlu-----ProProArgGluVal 459  
 QY 1265 GAGAGTGGTGTGAGAGAGGGGCGGACTTGGAGAGACCTGTGTGATCAAAAAGCTG 1324  
 |||||  
 Db 460 GlnProGlyGlnAlaArgGln---ProSerGlnGlnIleuLeuPheArgGlnGlnAlaLeu 478  
 QY 1325 TTTTCAGATGGCCCAACCGCTTTCAGAGGTACAGAGGTACCAAGGCGCCCTCAGCGTGGCC 1384  
 |||||  
 Db 479 LeuLeuGlnGlnGlnAlaGlnIleHisGlnLeuArgAsnTyrGlnAlaSerMetGlnAlaAla 498  
 QY 1385 ACTGTGGCC-----CACCAAGCCTTGGGCGCTTACCACTCCCTGCT 1429  
 |||||  
 Db 499 GlyIleProValSerPheGlyGlyHisArgProLeuSerArgAlaGlnSerSerProAla 518  
 QY 1430 GCC-----CCTGGGGGCAATGAGAGACCCCGCAGACCAACCGCTGACAGCCTTTC 1480  
 |||||  
 Db 519 SerIleThrPheProIleSerValGlnGlnProProThrLysPro-----ArgPhe 535  
 QY 1481 ACCCAAGATGTGTCTACAGACCTTCATGTCTAAGACCAAGCAGATGTGGGGAGACCA 1540  
 |||||  
 Db 536 ThrThrGlyLeuValTyrAspThrLeuMetLeuLysHisGlnCysThrCysGlySerSer 555

QY 1541 CACGTCGACCTGAGCATGTGGCGGAGTCCAGAGCATCTGTCCCGGCTGAGAGACA 1600  
 |||||  
 Db 556 SerSerHisProGlnHisAlaGlyArgIleGlnSerIleTyrSerArgLeuGlnGluThr 575  
 QY 1601 GGCCTGTCTTACCAAGTGGGAGCGGAGTCCGAGGTGCGAAAGCCAGCTAGATGATCCAG 1660  
 |||||  
 Db 576 GlyLeuArgGlyLysCysGluCysIleArgGlyThrGlySalThrLeuGlnGlnGln 595  
 QY 1661 ACAGTGCACCTGAAATPACACACCTGCTATATGGAGCAGCTCCCTCAACCGGCAAGAG 1720  
 |||||  
 Db 596 ThrValHisSerGlnAlaHisThrLeuLeuTyrGlyThrAsnProLeuAsnArgGlnLys 615  
 QY 1721 CTAGCAGACAGAAAGTGTGCTGCTCCCATGCGCCAGAAAGATGATGCTGTGCTGCTGT 1780  
 |||||  
 Db 616 LeuAspSerLysLysLeuLeuGlySerLeuAla---SerValPheValArgLeuProCys 634  
 QY 1781 GGGGCGCATCGGGGTGGACAGTGAACCCGTGTGAGATGAGATGACATCTCCAGTCTGTG 1840  
 |||||  
 Db 635 GlyGlyValGlyValAlaAspSerAspThrIleTyrAsnGlnValHisSerAlaGlyAlaAla 654  
 QY 1841 CGCATGTGAGTGGGCTGCTGCTGAGCTGAGCTGCTTCAAGTGGCTGCGAGAGACTCAAG 1900  
 |||||  
 Db 655 ArgLeuAlaValGlyCysValAlaGlyLeuValAlaThrGlyGlyLeuLys 674  
 QY 1901 AATGATTTGGCATCATCCGCGCCCGCAGAGACACCGCGGAGAAATCCACAGCCATGGGA 1960  
 |||||  
 Db 675 AsnGlyPheAlaValAlaArgProProGlyHisHisAlaGlnGlnSerThrProMetGly 694  
 QY 1961 TTTGCTTTCTTCAACTGTGTAGCCATCACGCGAAACTCTTACAGACAGAGTTGACGTG 2020  
 |||||  
 Db 695 PheCysTyrPheAsnSerValAlaValAlaLysLeuLeuGlnGlnArgLeuSerVal 714  
 QY 2021 GCGAGGTCCTCATCGTGGAGTGGAGATCACCATGGCATGGCAGCCAGCAGAGCGTTC 2080  
 |||||  
 Db 715 SerLysIleLeuIleValAspTyrAspValHisHisGlyAsnGlyThrGlnIleAlaPhe 734  
 QY 2081 TACAATGACCCCTCTGTGCTTATCATCTCTGTGATCGCTATGACAGGAACTTCTT 2140  
 |||||  
 Db 735 TyrSerAspProSerValLeuTyrMetSerLeuHisArgTyrAspArgGlyAsnPhePhe 754  
 QY 2141 CCAGGCTCTGGGGCTCTGAAAGAGTTGGTGGAGCAGCAGCGCTGGGTACAAATGTGAC 2200  
 |||||  
 Db 755 ProGlySerGlyAlaProAspGluValGlyThrGlyProGlyValGlyPheAsnValAsn 774  
 QY 2201 GTGGCAGGACGAGGAGGTGTGAGCCCCCATTTGAGAGCGTGAAGTCCCTTACAGCCCTTC 2260  
 |||||  
 Db 775 MetAlaPheThrGlyGlyLeuAspProProMetLysPheAlaGluTyrLeuAlaAlaPhe 794  
 QY 2261 AGCAGAGTGTGATGCCCATTTGCCAGAGTTCACCTGATGTGTGCTTGTCTCCGCC 2320  
 |||||  
 Db 795 ArgThrValValMetProIleAlaSerGlnPheAlaProAspValValLeuValSerSer 814  
 QY 2321 GGGTTGATGCTGTGAAGGACATGTGTCTCTGTGGTGGCTACTGTTCACCGCCAGA 2380  
 |||||  
 Db 815 GlyPheAspAlaValGlnGlnHisProThrProLeuGlyGlyTyrAsnLeuSerAlaArg 834  
 QY 2381 TGTTTGGCCATTCAGCAGCAGCTGATGACCCCTGGAGAGGCGCGGCGTGTGCTGCC 2440  
 |||||  
 Db 835 CysPheGlyTyrLeuThrLysGlnLeuMetGlyLeuAlaGlyGlyAlaLeuAla 854  
 QY 2441 CTGAGAGGAGGCGCATGACTTTCAGCGCATCTGTATGCTGTGAGCTGTGTCTCGGCT 2500  
 |||||  
 Db 855 LeuGlnGlyGlnHisAspLeuThrAlaIleCysAspAlaSerGlnValLysValSerAla 874  
 QY 2501 CTGCTCAGTGAAGCTGTGACGCTTGGATGAGCAGCTTGGAGCGAAAGCCCAACATC 2560  
 |||||  
 Db 875 LeuLeuGlnAsnGlnLeuAspProLeuProGlnLysValLeuGlnGlnArgProAsnAla 894  
 QY 2561 AACGCAGTGGCGCAGCTAGAGAAAGTCAATGATCCAGAGCAACAGCTGAGTGTG 2620  
 |||||  
 Db 895 AsnAlaValArgSerMetGlnLysValMetGlnIleHisSerLysTyrTTPArgCysLeu 914  
 QY 2621 CAGAAAGTTCCGCGCTGTGCTGGCGCGGTCCCTGCGAGAGGCGCCAAAGCAGGTGAGACGAA 2680

```
Db 915 GlnArGThrThrSerThrAlaGlyArGSerLeuIleGluAlaGlnThrCysGluAsnGlu 934
QY 2681 GAAGCCGAAT-GTGAAGCCATGGCCCTTGCTGTGGGGGCGGACAGAGCCCAAGCT 2739
Db 935 GluAlaGluThrThrAlaMetAlaSerLeuSerValGlyValLysProAlaGlu--- 953
QY 2740 GCGGACGCCCGGGAACACAGCCCGAGCCGCGAGAGAGCCCATGAGACAGAGCCCTGCC 2799
Db 954 -----LysArGProAspGluGluProMetGluGluGluProPro 966
QY 2800 CTG 2802
Db 967 Leu 967
RESULT 6
AAU99659 standard; Protein: 967 AA.
ID AAU99659 standard; Protein: 967 AA.
XX
XX AAU99659;
XX
XX 07-OCT-2002 (first entry)
XX
XX Human Histone deacetylase isoform 4.
XX
XX Human; histone deacetylase; HDAC-4; cancer; cytostatic;
XX antisense; tumour suppressor; cell proliferation; tumour;
XX programmed cell death; necrotic cell death.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 72
XX FT Misc-difference /note- "Encoded by CCG"
XX FT Misc-difference 78
XX FT Misc-difference /note- "Encoded by ACT"
XX FT Misc-difference 80..81
XX FT Misc-difference /note- "Encoded by AGAGAC"
XX FT Misc-difference 135
XX FT Misc-difference /note- "Encoded by TCA"
XX FT Misc-difference 194..196
XX FT Misc-difference /note- "Encoded by GCGTGGAGS"
XX FT Misc-difference 318
XX FT Misc-difference /note- "Encoded by GTA"
XX FT Misc-difference 574
XX FT Misc-difference /note- "Encoded by AAG"
XX FT Misc-difference 701
XX FT Misc-difference /note- "Encoded by GCG"
XX FT Misc-difference 775
XX FT Misc-difference /note- "Encoded by ACG"
XX FT Misc-difference 812
XX FT Misc-difference /note- "Encoded by GTC"
XX FT Misc-difference 948
XX FT Misc-difference /note- "Encoded by GAC"
XX
XX US2002061860-A1.
XX
XX 23-MAY-2002.
XX
XX 06-AUG-2001; 2001US-0817913.
XX
XX 24-MAR-2000; 2000US-192157P.
XX
XX (LIZZ/) LI Z.
XX PA (BONF/) BONFILS C.
XX PA (BEST/) BESTERMAN J.
XX
XX LI Z, Bonfils C, Besterman J;
XX
XX WPI: 2002-507650/54.
XX DR N-PSDB; ABK87718.
XX
XX Agent that specifically inhibits an isoform of histone deacetylase.
```

```
PT useful for treating cancer and other cell proliferative diseases,
PT preferably comprises an antisense oligonucleotide
XX
XX Disclosure; Fig 4A; 60pp; English.
XX
XX The invention relates to an agent that inhibits an isoform of histone
XX deacetylase (HDAC-1 to HDAC-8) but not all isoforms, e.g. an
XX antisense oligonucleotide. Also included are inhibiting an HDAC isoform
XX in a cell by treatment with the agent, identifying an HDAC isoform that
XX is required for induction of cell proliferation or differentiation and
XX inhibiting cell proliferation by treatment with two antisense
XX oligonucleotides or small molecules that inhibit a specific HDAC
XX isoform, or antisense oligonucleotide or small molecules that inhibit
XX DNA methyltransferase. The agent therefore acts as a tumour suppressor.
XX The agents are used to treat diseases of cell proliferation and
XX differentiation (e.g. cancer and tumours), by inducing growth retardation,
XX growth arrest or programmed/necrotic cell death, specifically neoplastic
XX cell proliferation in humans. The agents are selective for particular
XX isoforms, compared to known inhibitors which are not selective.
XX The present sequence represents the HDAC-4 isoform.
XX
XX SQ Sequence 967 AA;
XX
XX Alignment Scores:
XX Pred. No.: 1,08e-172 Length: 967
XX Score: 2685.00 Matches: 567
XX Percent Similarity: 71.28% Conservative: 118
XX Best Local Similarity: 59.00% Mismatches: 196
XX Query Match: 50.22% Indels: 81
XX Db: 23 Gaps: 23
XX
XX US-09-502-945-2 (1-2885) x AAU99659 (1-967)
QY 2 GAATTCCTCTGTGCGAAGTCAAGAGCCGACACAGCGGCGCTCAACATTCCTCCCA 61
Db 60 GluPheValIleuAsnLysLysLysAlaLeuAlaHisProAsnLeuAsnHisCysIleSer 79
QY 62 CAGCACCCCAATGCTGG-----GGAGCCCAAGCTCTTTGGACAGATGCCCT 115
Db 80 SerCysProArGlyTrpTyrGlyLysThrGlnHisSerSerLeuAspGlnSerSerPro 99
QY 116 CCCGAGAGCGGCCCGCCCGGAGCGCTCCCTCCCAAGTGGCTTCGCGGCGCTCA 175
Db 100 ProGlnSer-----GlyValSerThrSerTyrAsnHisProValLeuGlyMetLys 116
QY 176 GACAGTCGAGAGCACTCCCTCCGCAAAACAGCCTTGAACCACTTGAAGTGGCT 235
Db 117 AspaIaLysAspaSerPheProLeuArGlyThrAlaSerGluProAsnLeuLysLeuArg 136
QY 236 TCAGGCTAAACAGAGGTGGCTGAGCGAGAGACAGTCCCTCTCGTCGCAAGAT 295
Db 137 SerArgLeuLysGlnLysValAlaGluArgArgSerSerProLeuIleArgLysAsp 156
QY 296 GGGACTGTATTAGACCTTTAAGAGAGAGAGCTGTGGATCATCAGGCGCGGCTGG 355
Db 157 GlyProValValThrAlaLeuLysLysArgProLeuAspValThr----- 171
QY 356 GCGTCGTCGCTGTGTAACAGCGACCCGCTCCGCGCCAGCTCTCC---AACAGCTCC 412
Db 172 ---AspSerAlaCysSerSerAlaProGlySerGlyProSerSerProAsnAsnSerSer 190
QY 413 CACAGCACCATCGCTGAGAAATGGCTTTACTGGCTCAGTCCCAACATCCCACTGAGATG 472
Db 191 GlySerValSerAlaGluAsnGlyLeuAlaProAlaValProSerIleProAlaIuThr 210
QY 473 CTCCTCAGACACCGAGCCCTCCCTGTGACAGAGCTCCCAACAGTTCAGCTCAGACG 532
Db 211 SerLeuAlaHisArgLeuValAlaArgGluGluSerAlaAlaProLeuProLeuLysThr 230
QY 533 TCTCCTTCTGCGCAACATCTCCCTAGCGCTGAGCGGCGACGATCAGTCAACCAACTCA 592
Db 231 SerProSerLeuProAsnIleThrLeuGluLeuProAlaIthr----- 244
```



QY 593 CACSTCACTGCTCCCGAAGCTGTGCACA-----CAGCAGAGAGCGGAGAGCAGAGCC 646  
 Db 245 -----GlyProSerAlaGlyThrAlaGlyLeuGlnAspThrGlnArgLeuThr 260  
 QY 647 CTCGAGTCCCTGGCGGAG-----GATGGCAGCGTACCGGCAAGTTCATG 691  
 Db 261 LeuProAlaLeuGlnGlnArgLeuSerLeuPheProGlyThrHisLeuThrProTyrLeu 280  
 QY 692 AGCAGATCCTCTATCTCTGGCTGGCTGGCGCTGAGCTGAGAGGGGAGAGAGGCCCC 751  
 Db 281 SerThrSer-----ProLeuGlnAArgAspGlyAla 291  
 QY 752 CACGGGAGCTCCCTGCTGCTGACAGATGCTGTGTGAGAGAGCGCCGACAGAGC 811  
 Db 292 ---AlaHisSerProLeuLeuGlnHisMetValLeuLeuGlnGlnProProAlaGlnAla 310  
 QY 812 ACCCTCAAT-----GCTGTGCCACTCCAGGGGAGTCCCACTAGTACGGGT 859  
 Db 311 ProLeuValThrGlyLeuGlyAlaLeuProLeuHisAlaGlnSer---LeuValGlyAla 329  
 QY 860 GAACGTGGGCGCAGCAGATCGGAGGTAGCAGAGCTCCGCGGCGCATCGGCGCTGAGC 919  
 Db 330 AspArgValSerProSer-----IleHisLeuAArgGlnHisAlaArgProLeuGly 346  
 QY 920 CGCAGTCACTCTCAGCGCTGCGCAGAGATCCCGAGCGCTGACAGAGCTGTGATGCAA 979  
 Db 347 ArgThrGlnSerAlaProLeuProGlnAsnAlaGlnAlaLeuGlnHisLeuValIleGln 366  
 QY 980 CAAAGACACCAAGCAAGTTCTCTGGAGAGCAGAG-----CAGCAGCAGCTACAG 1027  
 Db 367 GlnGlnHisGlnGlnPheLeuGlnGlnLysGlnHisGlnGlnPheGlnGlnGlnGln 386  
 QY 1028 CTGGGAGAGATCTCTCAGCAGACAGAGGAGAGCTGCCAGAGCGCCACAGCCAGCTGAG 1087  
 Db 387 MetAsnLysIleIleIerProLysProSerGlnProAlaArgGlnProGlnSerHisProGln 406  
 QY 1088 GAGACAGAGAGAGAGCTGACAGCAGCAGCAGAGAGCTTGTGGGAGAGAGCCCTG--- 1144  
 Db 407 GlnThrGlnGlnGlnGlnLeuArgGlnHisGln---AlaLeuLeuAspGlnProTyrLeuAsp 425  
 QY 1145 ACATGCCCCGGGAGGCTCCACAGAGAGTACAGACACAGGAAGACTGTGAGAGAGAG 1204  
 Db 426 ArgLeuProGlnLysGlnAlaHisAlaGlnAlaGlyValGln---ValLysGlnGln 444  
 QY 1205 GACGAGAGAGAGATGGGAGGAGAGAGAGATTCATCCAGCTTAAGAGAGAGAGGCG 1264  
 Db 445 ProIleGlnSerAspGlnGlnGlnAlaGln-----ProProArgGlnVal 459  
 QY 1265 GAGAGTGTGCTGAGAGAGGCGCCGACTTGGAGAGAGCTGTGCTGATCAAAAGACTG 1324  
 Db 460 GlnProGlnArgGln---ProSerGlnGlnGlnLeuLeuPheAlaGlnGlnAlaLeu 478  
 QY 1325 TTCTCAGATGCCCAACCGCTGCAACTTTTGAGGTGTACCAAGCGCCCTCAGCCTGGCC 1384  
 Db 479 LeuLeuGlnGlnGlnArgGlnHisGlnLeuArgAsnTyrGlnAlaSerMetGlnAla 498  
 QY 1385 ACTGTGCC-----CACCAAGCCTGGGCGCTACCCAGTACTCCCTGCT 1429  
 Db 499 GlnIleProValSerPheGlnGlyLysArgProLeuSerAlaGlnSerSerProAla 518  
 QY 1430 GCC-----CTGGGGGCGATGAAGACCCCGCAGACCAACCGCTCAAGCAGCTTTC 1480  
 Db 519 SerIleThrPheProValSerValGlnGlnProProThrLysPro-----ArgPhe 535  
 QY 1481 ACCAGAGTGTGTCTACGACAGCTTCATGCTTAAGACAGAGTGCATGTGCGGGAGACA 1540  
 Db 536 ThrThrGlyLeuValTyrAspThrLeuMetLeuLysHisGlnLysThrCysGlySerSer 555  
 QY 1541 CACGTGACCTTGAGTGTGCGCGGAGATCCAGAGCATGTGTGCTCCGCGCTGACAGAGACA 1600  
 Db 556 SerSerHisProGlnHisAlaGlyArgIleGlnSerIleThrSerArgLeuGlnGlnThr 575  
 QY 1601 GGCTGTGTTAGCAAGTGCAGAGCGGATCCGAGGTGCAAAAGCCAGCTAGATGAGATCCAG 1660

Db 576 GlnLeuArgGlyLysCysGlnCysIleArgGlyArgLysAlaThrLeuGlnGln 595  
 QY 1661 ACACTGACCTCTGAATACACACCCCTGCTATGGAGACCACTCCCTCAACCGGACAG 1720  
 Db 596 ThrValHisSerGlnAlaHisThrLeuLeuTyrGlyThrAsnProLeuAsnArgGlnLys 615  
 QY 1721 CTAGACAGCAAGAGTGTGCTGGCTCCATCAGCCAGAAAGATGTGCTGCTGCTCTGT 1780  
 Db 616 LeuAspSerLysLysLeuLeuGlySerLeuAla---SerValPheValArgLeuProCys 634  
 QY 1781 GGGGCGATCGGGGTGACAGAGTACAGACCCGTGGAGATGAGATGACCTCTCAGTGGTGTG 1840  
 Db 635 GlyGlyValGlyValAspSerAspThrIleTyrAsnGlnValHisSerAlaGlyAlaAla 654  
 QY 1841 CGCATGCACTGGGCTGCTGCTGCTGAGCTGAGCTGCAAGGTGCTGCAAGAGACTCAAG 1900  
 Db 655 ArgLeuAlaValGlyCysValValGlnLeuValPheLysValAlaThrGlyGlnLeuLys 674  
 QY 1901 AATGAGTTTGGCATCCCGGCGCCGAGAGACACCGCGGAGAAATCCACAGCCATGGGA 1960  
 Db 675 AsnGlyPheAlaValAlaArgProProGlyHisHisAlaGlnGlnGlnThrProMetGly 694  
 QY 1961 TTCTGCTTCTTCAACTGTGTAGCCATCAGCCGAAAGCTCTTACAGAGAAATTGAAGCTG 2020  
 Db 695 PheCysTyrPheAsnSerValAlaValAlaAlaLysLeuLeuGlnGlnArgLeuSerVal 714  
 QY 2021 GCGAGGTCTCTCATCTGAGACTGGGAGACATTCACATGGCAATGCCACCCAGAGCGCTTC 2080  
 Db 715 SerLysIleLeuValAlaAspTyrAspValHisHisGlyAsnGlnGlnGlnAlaPhe 734  
 QY 2081 TACATGACCCCTGTGCTGTACATCTCTGTGATGCGTATGACAGCGGAACTTCTT 2140  
 Db 735 TyrSerAspProSerValLeuTyrMetSerLeuHisArgTyrAspAspGlyAsnPhePhe 754  
 QY 2141 CCAAGCTCTGGGCTCTGTAAGAGATTGTGTGAGAGACAGAGCGGTGGGATCAATGTGAAC 2200  
 Db 755 ProLysSerGlyAlaProAspGlnValGlyThrGlyProGlyValGlyPheAsnValAsn 774  
 QY 2201 GTGGCAGTACAGAGAGTGTGTGAGACCCCGCCATTTGAGAGCGAGTACCTTACAGCTTC 2260  
 Db 775 MetAlaPheThrGlyGlyLeuAspProProMetLysAspAlaGlyTyrLeuAlaAlaPhe 794  
 QY 2261 AGCAGAGTGTGATGCCCATTTGCCACAGAGTTCACCTGTATGTGGTCTAGTCCGCC 2320  
 Db 795 ArgThrValAlaMetProIleAlaSerGlnPheAlaProAspValAlaLeuAlaSerSer 814  
 QY 2321 GGGTTTGATGCTGTTGAAGAGCATCTGTCTCTGGTGGCTACTGTCTACCGCCAGA 2380  
 Db 815 GlyPheAspAlaValGlnGlnLysProThrProLeuGlnGlyTyrAsnLeuSerAlaArg 834  
 QY 2381 TGTTTTGGCCACTTGACCAAGCAGCTGATGACCTGCGAGGGGCGGGTGTGCTGGCC 2440  
 Db 835 CysPheGlnTyrLeuThrLysGlnLeuMetGlyLeuAlaGlyArgIleValLeuAla 854  
 QY 2441 CTGGAGGAGGCGCATGACTTGACCGCCATCGTATGATGCTGTGAAGCTGTGTGCTGGCT 2500  
 Db 855 LeuGlnGlyGlnHisAspLeuThrAlaIleCysAspAlaSerGlnAlaCysValSerAla 874  
 QY 2501 CTGCTCAGTGTAAAGCTGACAGCTCTTGATGAGAGCACTTCTTGCAGCAAAAGCCCAACATC 2560  
 Db 875 LeuLeuGlnLysGlnGlnLeuAspProLeuProGlnLysValLeuGlnGlnArgProAsnAla 894  
 QY 2561 AACGCACTGGCCACGCTAGAGAAAGTATCAGATCCAGAGCAAAACATGAGCTGTGTG 2620  
 Db 895 AsnAlaValAlaArgSerMetGlnLysValMetGlnIleHisSerLysTyrTrpArgCysLeu 914  
 QY 2621 CAGAAGTTCCGCGTGTGCTGCGCGGCTCCCTGCGAGAGGCGCCAGAGAGTGTGAGCCGA 2680  
 Db 915 GlnArgThrThrSerThrAlaGlyArgSerLeuLeuGlnAlaGlnThrCysGlnAsnGln 934  
 QY 2681 GAAGCCGAAAT--GTGAAGCGCATGCGCTTGTGCTGTGGGGGCGCAAGGCCAAGCT 2739

D	b	935	GluHlaGIuThrValTThrAlaMetAlaSerLeuSerValGlyValLysProAlaGlu---	953
O	y	2740	GGCGACGCCGGGAACAACAGCCCCAGCCGGCAGAGAGAGCCCATGTGACGACGAGCTGC	2799
D	b	954	-----LysArgProAspIuclutInuPrometGluInuGuPuroPro	966
O	y	2800	CTG 2802	
D	b	967	Leu 967	
R	E	S	RESULT 7	
I	D	AAM93526		
I	D	AAM93526	standard; Protein; 417 AA.	
A	C	AAM93526;		
X	X	06-NOV-2001	(first entry)	
D	T	Human polypeptide,	SEQ ID NO: 3261.	
X	X	Human; full length cDNA; cDNA synthesis; oligo-capping.		
X	X	Homo sapiens.		
X	X	EP1130094-A2.		
X	X	05-SEP-2001.		
P	E	07-JUL-2000;	2000EP-0114089.	
X	X	08-JUL-1999;	99JP-0194486.	
P	R	11-JAN-2000;	2000JP-0118774.	
P	R	02-MAY-2000;	2000JP-0183765.	
P	A	(HELI-) HELIX RES INST.		
X	X	Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;		
P	I	Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;		
X	X	WPI: 2001-524255/58.		
D	R	N-PSDB; AAK94456.		
X	X	830 Primers useful for synthesizing full length cDNA clones and their		
P	T	use in genetic manipulation -		
X	X	Claim 8; SEQ ID NO 3261; 1380bp + sequence listing; English.		
P	S	The invention relates to primers for synthesizing full length cDNA		
X	X	clones. 830 cDNA molecules encoding a human protein have been		
C	C	isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA		
C	C	molecules have been determined. Primers for synthesizing the full length		
C	C	cDNA are useful for clarifying the function of the protein encoded by		
C	C	the cDNA. The full length clones were obtained by construction of full		
C	C	length enriched cDNA libraries that were synthesised by the oligo-capping		
C	C	method. The primers enable the production of the full length cDNA easily		
C	C	without any special methods. The present sequence is a polypeptide		
C	C	encoded by a full length human cDNA of the invention.		
C	C	Note: The sequence data for this patent did not form part of the printed		
C	C	specification, but was obtained in CD-ROM format directly from EPO.		
X	X	Sequence 417 AA;		
A	L	ignment Scores:		
P	r	ed. No.: 3,62e-135	Length: 417	
S	c	ore: 2127.00	Matches: 413	
P	e	cent Similarity: 99.28%	Conservative: 2	
B	e	st local Similarity: 98.80%	Mismatches: 2	
Q	u	ery Match: 39.79%	Indels: 1	
D	B	: 22	Gaps: 0	
U	S	-09-502-945-2 (1-2885) x AAM93526 (1-417)		
O	y	470 ATGTCCTCCATGACGACGAGCCCTCCCTCTGTGACAGCTCCGCCCAACGAGTTCA	529	

Db 361 ProGluHisAlaGlyArgIleGlnSerIleTrpSerArgLeuGlnIuThrGlyLeuLeu 380  
 QY 1610 AGCAAGTGCAGCGGATCCGAGTCCGCAAGCCAGCTAGATGAGATCAGACAGTCAC 1669  
 |||||  
 Db 381 SerIleGlySerGlyLeuArgGlyArgGlyAlaThrIleuAspGlu-ThrArgGlnCysTh 400  
 QY 1670 TCTGATATACCAACCCCTGCTCTATGGAGCCAGTCCCTCAACGGCAGAAAC 1721  
 |||||  
 Db 400 rIeuAsnThrThrProCysSerMetGlyProValProSerThrGlyArgSer 417

RESULT 8

AAM78891

ID AAM78891 standard; Protein; 1008 AA.

XX AAM78891;

DT 06-NOV-2001 (first entry)

DE Human protein SEQ ID NO 1553.

XX Human protein SEQ ID NO 1553.

XX Human protein SEQ ID NO 1553.

KW Human: cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;

KW nervous system disorder; arthritis; inflammation.

XX Homo sapiens.

PN MO200157190-A2.

XX 09-AUG-2001.

PF 05-FEB-2001; 2001MO-US04098.

XX 03-FEB-2000; 2000US-0496914.

PR 27-APR-2000; 2000US-0560875.

PR 20-JUN-2000; 2000US-0598075.

PR 19-JUL-2000; 2000US-0620325.

PR 01-SEP-2000; 2000US-0654936.

PR 15-SEP-2000; 2000US-0663561.

PR 20-OCT-2000; 2000US-0693325.

PR 30-NOV-2000; 2000US-0728422.

XX (HYSE-) HYSEQ INC.

XX (HYSE-) HYSEQ INC.

XX (HYSE-) HYSEQ INC.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;

PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI: 2001-476283/51.

DR N-PSDB; AAK52024.

PT Nucleic acids encoding polypeptides with cytokine-like activities,

PT useful in diagnosis and gene therapy -

XX Claim 20; Page 3870-3872; 6221pp; English.

PS The invention relates to polynucleotides (AAK51456-AAK53435) and the

XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to

CC cytokine, cell proliferation or cell differentiation or which may induce

CC production of other cytokines in other cell populations. The

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,

CC e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and

CC activity/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation.

CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666

CC (AAM80020) are omitted as the relevant pages from the sequence listing

CC were missing at the time of publication.

XX Sequence 1008 AA;

## Alignment Scores:

Pred. No.: 3,96e-115

Score: 1831.50

Percent Similarity: 56.03%

Best Local Similarity: 43.35%

Query Match: 34.26%

DB: 22

US-09-502-945-2 (1-2885) x AAM78891 (1-1008)

QY 11 TTGTCGAAGTCAAGAGAGCCACACAGGCGGCTCACCATTCCCTCCACAGACACC 70  
 ||| : : : : ||| : : : :  
 Db 157 LeuGluArgThrValHisProAsnSerProGlyIleProTyArgThrLeuGluProLeu 176  
 QY 71 AAMTGCCTGGAGGCCACCAATGCTTCTTGACACAGTTCCTCCACAGGCGGCC 130  
 : : : : ||||| : : : : |||||  
 Db 177 GluThrGluGlyAlaThrArgSerMetLeuSerSerPheLeuProProValProSerLeu 196  
 QY 131 CTTGGAGCGGCTCCCTCTCAAACTGCTTTCCTGGGCTTACGACAGTGCAGAGAC 190  
 ||| ||||| : : : :  
 Db 197 ProSerAspProProGluHis----- 203  
 QY 191 TTCCCTCCCTCGCAAAACAGCCTCTGAACCCACTGAAAGTGGCTTCAAGCTTAAACAG 250  
 ||||| ||||| ||||| ||||| ||||| : : : |||||  
 Db 204 PheProLeuArgIleThrValSerGluProAsnLeuIleuIleuIleuIleuIleuIleuIleu 222  
 QY 251 AAGTGGCTGACGGGAGCAACAGATCCCTCTGCTGCGCAAGGATGGACTTTATTAGC 310  
 ||| ||||| ||||| : : : ||||| ||||| : : : :  
 Db 223 LysSerLeuGluArgValGlyAsnProLeuLeuArgLysGluSerAlaPro-----Pro 240  
 QY 311 ACCTTTAAAGAGAGAGCTGTGATACAGATCAGGCGGCGGCTGCGGCTCGTGTGT 370  
 : : : : : : : : ||| : : : : : : : :  
 Db 241 SerLeuArgArgArgProAlaGluThrLeuGlyAspSerSerProSerSer----- 258  
 QY 371 AACAGCGCACCGGCTCGGCGCCAGCTCTCCCAACAGCTCCACAGACACCATTCGTGAG 430  
 ||| ||| ||||| ||||| ||||| |||||  
 Db 259 ---SerThrProAlaSerGlyCysSerSerProAsnAspSerGluHis----- 273  
 QY 431 AATGGCTTTACTGCTACGTACGTCGTCACATCCCAACATCCCAACATCCCAACATCCCAAC 490  
 ||| ||| : : : : : : : : ||| |||  
 Db 274 -----GlyProAsnProIleLeuGlySerIleuAlaLeuGluGlyIleuArgLeu 289  
 QY 491 CTCCTCTGAGACAGTCCGCCAACCAACAGTTCAGCTTACAGCTCTCTTCTGCCCAAC 550  
 : : : : : : : : ||||| : : : : : : : :  
 Db 290 ArgLeuGlnIuThrSerValAlaProPheAlaLeuProThrValSerLeuLeuProAla 309  
 QY 551 ATCTCCCTAGGGCTGACAGGCGACGGTCACTGACCAACTCAGCACTGACTGCTCCCGC 610  
 ||| : : : ||||| |||  
 Db 310 IleThrLeuGlyLeuProAla----- 316  
 QY 611 AAGCTGTGCACAGCAGGAGCGCCGAGAGAGCCCTCCAGTCCCTGCGGCGAGGTTGCG 670  
 : : : : : : : : ||||| : : : : : : : :  
 Db 317 -----ProAlaArgAlaAspSerAspArgArgThrHisProThrLeuGlyProArgGly 334  
 QY 671 ACGGTGACGGCAAGTTCATGAGCAGACATCTCTATTCTGGTGGCTGCGGTGAGTGC 730  
 : : : |||| : : : : : : : :  
 Db 335 ProIleLeuGlySerProHisThrProLeuPheLeuProHis-----GlyLeuGlu 351  
 QY 731 CTGAGAGGCGAGCGGAGCCGCCACAGGCGATCCCTCGTGCAGCATGTGCTGTGCTG 790  
 ||| : : : : ||| ||||| : : : |||||  
 Db 352 ProGluAlaGlyThr-----LeuProSerArgLeuGlnProIleLeuLeuLeu 368  
 QY 791 GAGCAGGCGCGGAGCAGCAGACCCATTCGTGTGTGCATCCACAGGCGAGTCCCACTA 850  
 : : : : : : : : ||||| |||  
 Db 369 AspProSerGlySerHisAlaProLeuLeuThrValProIleLeuGlyProLeuProPhe 388  
 QY 851 GTGACGGGTGAAGCTGGGCGACAGCATCGACAGGATAGGATCCCGCGGATCGG 910  
 : : : : : : : : ||| |||  
 Db 389 HisPheAlaIleuSerLeuMetThrThrGluArgLeuSerIleSer---GlyLeuHisTrp 407  
 QY 911 CCCCTGAGCGGCACTAGTCTCAGCGCTGCGCGAGAGTCCAGAGCC----- 958  
 ||||| ||||| ||||| ||||| ||||| |||||  
 Db 408 ProLeuSerArgThrArgSerGluProLeuProProSerAlaThrAlaProProPro 427

0Y	959	-----CTGACGAGCCTGGTCATGTCACAAACAGACAGACAGGTCCTG	1000
0Y	428	GlyProMetGlnProAlaGlyLeuGlnLeu-----LysThrHisValGlnValIle	444
0Y	1001	GAGAAAGCAGAAAGCAGACAGACTACAGACTGGGCGAAGATCCTCAACCAAGACAGGAGCTG	1060
Db	445	LysArgSer-----AlaLysProSerGlnLys	453
0Y	1061	CCCAAGGACGCCACCACCACCTCGAGAGACACAGAGAGAGACTGACGAGACGACGAGAG	1120
Db	454	ProArg-----	455
0Y	1121	GTCCTTGCGGGGAGAGAGCCCTGACCATGCCCGGGAGAGGCTCCACAGAGATGAGAGC	1180
Db	455	-----	455
0Y	1181	ACACAGGAAGACCTGGAGAGAGAGACGAGAGAAAGAGATGGGAGAGAGAGAGAGATTC	1240
Db	456	LeuArgGlnIleProSerAlaGlnAspLeuGlnThrAspGly-----	469
0Y	1241	ATCCAGGTTAAGGACGAGAGAGGCGAGAGTGGT-----GCTGAGAGAGGCGCCGACCTG	1294
Db	470	-----GlyGlyProGlyGlnValValAlaAspAspGlyLeuGlnHis	482
0Y	1295	GAGGAGCCTGGCTGGATACAAAAAACTGTCTCAAGATGCCAACCCGTGCACCTTTG	1354
Db	483	ArgGluLeuGlnHisGlyGln-----ProGlnAlaArgGlyProAlaProLeu	498
0Y	1355	CAGGTGTACCAAGGCC-----	1372
Db	499	Gln-----GlnHisProGlnValLeuLeuTrpGlnGlnIleArgLeuAlaGlyArgLeu	516
0Y	1373	-----CTCAGGCTGGCCACATGTGTGCCACCAAGCC	1402
Db	517	ProArgGlySerThrGlyAspThrValLeuLeuProLeuAlaGlnGlyHisArgPro	536
0Y	1403	CTGGGCGCTACCCCAATCCTCCCTCGCTGCGCCCTGGGGGCGATGAAGAACCC-----	1453
Db	537	LeuSerArgAlaGlnSerProAlaAlaProAlaSerLeuSerAlaProGluProAla	556
0Y	1454	-----CCAGACCAAAACCCGTACAGCAGCTC---TTCAACACA	1486
Db	557	SerGlnAlaArgValLeuSerSerSerGlnThrProAlaArgThrLeuProPheThrThr	576
0Y	1487	AGTGGGTCTACGACAGCTTCATGCTCTAAACACACAGTGCATGGCGGGAACACACAGCTG	1546
Db	577	GlyLeuIleIleTyrAspSerValMetLeuLysHisGlnCysSerCysGlyAspAsnSerArg	596
0Y	1547	CACCTGTGCATGTCTGGCCGCGATCCAGAGCATCTGGTCCCGGCTGCAGAGACAGACCTG	1606
Db	597	HisProGlnHisAlaGlyArgIleGlnInsertIleTrpSerArgLeuGlnGlnArgGlyLeu	616
0Y	1607	CTTACAGAGTCGAGCGGATCCGAGGTGCGAAAGCCACAGCTAGATGAGATCCAGACAGTG	1666
Db	617	ArgSerGlnCysGlnCysLeuArgGlyArgGlyAlaSerLeuGlnGlnLeuGlnInsertVal	636
0Y	1667	CACCTGAATCAACACCCTGTCTATAGGACACCAAGTCCCTCAACCGGACGAAGCTAAG	1726
Db	637	HisSerGlnArgHisValLeuLeuTyrGlyThrAsnProLeuSerArgLeuLysLeuAsp	656
0Y	1727	AGCAAGAGTTCGCGCTCCATCCAGACAGCAAAACATATGCTGTGGCTGTGGGAGGC	1786
Db	657	AsnGlyLysLeuAlaGlyLeuLeuAlaGlnArgMetPheValMetLeuProCysGlyIle	676
0Y	1787	ATCGGGGTGACACTGCACACCTGTGGATGAGATGACACTCCTCCAGTCTGTGGCATG	1846
Db	677	ValGlyValaSerThrAspThrIleTrpAsnGlnLeuHisSerSerAsnAlaIleArgTrp	696
0Y	1847	GCAAGGGGCTCCCTGCTGAGAGCTGGCTTTAAAGTGGCTGCGAGAGAGCTCAAAATGGA	1906
Db	697	AlaAlaGlnSerValIleThrAspLeuAlaPheLysValAlaSerArgGlnLeuLysAsnGly	716

[illegible]

KW nervous system disorder; arthritis; inflammation.  
 OS Homo sapiens.  
 XX W0200157190-A2.  
 XX 09-AUG-2001.  
 PD  
 XX 05-FEB-2001; 2001WO-US04098.  
 PE  
 XX 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 PR 20-JUN-2000; 2000US-0598075.  
 PR 19-JUL-2000; 2000US-0620325.  
 PR 01-SEP-2000; 2000US-0654936.  
 PR 15-SEP-2000; 2000US-0663561.  
 PR 20-OCT-2000; 2000US-0693325.  
 PR 30-NOV-2000; 2000US-0728422.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AU, Yang Y, Wejrtman T, Goodrich R;  
 XX  
 DR WPI; 2001-476283/51.  
 DR N-PSDB; AAK53008.  
 PT Nucleic acids encoding polypeptides with cytokine-like activities,  
 PT useful in diagnosis and gene therapy -  
 XX  
 XX Claim 20; Page 369-370; 6221pp; English.  
 PS  
 CC The invention relates to polynucleotides (AAK51456-AAK51435) and the  
 CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activity/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAM80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.  
 CC  
 XX  
 SQ Sequence 1020 AA;  
 Alignment Scores:  
 Pred. No.: 3,98e-115 Length: 1020  
 Score: 1831.50 Matches: 424  
 Percent Similarity: 56.03% Conservative: 124  
 Best Local Similarity: 43.35% Mismatches: 249  
 Query Match: 34.26% Indels: 182  
 DB: 22 Gaps: 22  
 US-09-502-945-2 (1-2885) x AAM79875 (1-1020)  
 QY 11 TTGTGCAAGTCAAGAGCCACACAGGCGGCTCAACATTCCTCCACAGACACCCC 70  
 DB 169 LenuLnuArgrThrValHISProAsnSerProGlyLeuProGlyThrLeuGluProLeu 188  
 QY 71 AAATGCGGGAGGACCCACATGCTTTTGGACAGAGTTCCCTCCACAGAGGCGCCC 130  
 DB 189 GlnThrGlnGlyAlaThrArgSerMetLeuSerSerPheLeuProProValProSerLeu 208  
 QY 131 CTTGGAGAGCTCCCTCTACAACTGCTTTGGCTGGCCCTTACACAGTGCAGAGCAGAC 190  
 DB 209 ProSerAspProProGluHis----- 215  
 QY 191 TTCCCTCGCGCAAAACAGGCTGTGAACCAACTTGAAGTGGCTTCAAGGCTAAACAG 250

DB 216 PheProLeuArgrLysThrValSerGluProAsnLeuLysLeuArgTyrLysProLys--- 234  
 QY 251 AAGTGGCTGAGCGGAGAGAGATGCCCTCCTCGCTGCAGAGATGGAGCTATTATAGC 310  
 DB 235 LysSerLeuGlnArgArgLysAsnProLeuLeuArgLysGlnSerAlaPro-----Pro 252  
 QY 311 ACCCTTAAGAGAGAGCTGTGAGATCATCAGAGTCCCGGCGCTGGGGGCTGCTCGTGTGT 370  
 DB 253 SerLeuArgrArgArgProAlaGlnThrLeuLysAspSerProSerSer----- 270  
 QY 371 AACAGCGCAGCCGCGCTCGGCGCCAGCTCTCCCAACAGCTCCACACACACATGCTGAG 430  
 DB 271 ---SerThrProAlaSerGlyCysSerSerProAsnAspSerGluHis----- 285  
 QY 431 AATGGCTTTACTGGCTCAGTCCCAACATCCCACTAGATGCTCCTCAGACCGAGCC 490  
 DB 286 -----GlyProAsnProIleLeuGlySerIleAlaLeuGlnArgLeu 301  
 QY 491 CTCCTCTGGACAGCTCCCGCCACACAGTTCAGCCTCTACACGTCCTCTCTGCGCCAC 550  
 DB 302 ArgLeuGlnGlnThrSerValAlaProPheAlaLeuProThrValSerLeuLeuProAla 321  
 QY 551 ATCTCCCTAGGGCTGACAGGCCAGCGTCATGTCACCACTCACTCACTGCTCCCGC 610  
 DB 322 IleThrLeuGlyLeuProAla----- 328  
 QY 611 AAGCTGTGCACACAGAGAGGCCGAGAGAGCGGCTCCAGTCCGCGGAGAGGTGC 670  
 DB 329 -----ProAlaArgAlaAspSerAspArgThrHisProThrLeuGlnProArgGly 346  
 QY 671 ACGGTACCGGGCAGTTCATGAGCACAATCCCTTCTGCTGCTCCGCTGGGGGTGGCA 730  
 DB 347 ProIleLeuGlySerProHisThrProLeuPheLeuProHis-----GlyLeuGlu 363  
 QY 731 CTGGAGGCGCAGCGGAGGCCCGCCAGGCGATGCTCCCTGCTGACAGATGCTGTTGCTG 790  
 DB 364 ProGluAlaGlyGlyThr-----LeuProSerArgLeuGlnProIleLeuLeu 380  
 QY 791 GAGAGCGCGCGGACGACAGAGACCCCTATTGCTGTGCTCCTCAGCCAGGCGAGTCCCACTA 850  
 DB 381 AspProSerGlySerHisAlaProLeuLeuThrValProGlyLeuGlyProLeuProPhe 400  
 QY 851 GTGAGGGGTGAAGCTGTGGCCACGACGATGGGACGATGAGGAGTCCCGGCGCATCGG 910  
 DB 401 HisPheAlaGlnSerLeuMetThrThrGlnArgLeuSerGlySer---GlyLeuHisTrp 419  
 QY 911 CCCCTGAGCGGACCTCAGTCCCTACCGCTGCCGACAGAGTCCCGAGGCC----- 958  
 DB 420 ProLeuSerArgThrArgSerGlnProLeuProProSerAlaThrAlaProProPro 439  
 QY 959 -----CTCCAGCAGCTGTGATGCAACAACAGCAGCCAGCTTCTG 1000  
 DB 440 GlyProMetGlnProArgLeuGlnGlnLeu-----LysThrHisValGlnValIle 456  
 QY 1001 GAGAGCAGAGAGAGCAGCAGAGCTACAGCTGGGCAAGATCTCAACCAAGAGAGGAGAGCTG 1060  
 DB 457 LysArgSer-----AlaLysProSerGluLys 465  
 QY 1061 CCCAGGACGCCACACACACACCTGAGAGAGACAGAGAGAGTGAAGAGAGAGAGAGAG 1120  
 DB 466 ProArg----- 467  
 QY 1121 GTCTTGCTGGGGAGGAGAGCCCTGACCATGCCCCGGAGAGGCTCCACAGAGATGAGAGC 1180  
 DB 467 ----- 467  
 QY 1181 ACACAGAGAGAGCTGAGAGAGAGAGAGAGAGATGGAGAGAGAGAGATGG 1240  
 DB 468 LeuArgGlnIleProSerAlaGlnAspLeuGlnThrAspGly----- 481  
 QY 1241 ATCCAGGTTAAGAGCAGAGAGAGAGAGTGTG-----GCTGAGAGAGGGCGCGACTTG 1294

Db 482 -----GlyIyProGIyGIValValAspAspGIyLeuGIuHis 494  
 QY 1295 GAGGAGCCTGCTGCTGATACAAAAAAGCTTTCTGATGCCCAAGCCTTG 1354  
 Db 495 ArgGIyLeuGIyHisGIyIn -----ProGIuAlaArgGIyProAlaProLeu 510  
 QY 1355 CAGGTGTACCAAGCGCC----- 1372  
 Db 511 Gln-----GlnHisProGIuValLeuLeuTrpGIuGlnArgLeuAlaGlyArgLeu 528  
 QY 1373 -----CTCAGCCTTGCCCACTTGCCCAACCAAGCC 1402  
 Db 529 ProArgGIySerThrGIyAspThrValLeuLeuProLeuAlaGlnGIyGlnHisArgPro 548  
 QY 1403 CTGGGCGCTACCAACCTCTCCCTGCTGCCCTGGGGCATGCAAGAACCC----- 1453  
 Db 549 LeuSerArgAlaGlnSerSerProAlaAlaProAlaSerLeuSerAlaProGIuProAla 568  
 QY 1454 -----CCAGAACACCCGTCAGACCTC---TTCACCCACA 1486  
 Db 569 SerGIuAlaArgValLeuSerSerSerGIuThrProAlaArgThrLeuProPheThrThr 588  
 QY 1487 AGTGTGCTTACGACACCTTCATGCTAAAGCAACCACTGATGGCGGAACACACAGCTG 1546  
 Db 589 GlyLeuIleTyrAspSerValIleMetLeuLysHisGlnCysSerGIyAspAsnSerArg 608  
 QY 1547 CACCCGTGACATGCTGGCGGATCCAGACATCTGCTCCGCTGCAGAGACAGCCTG 1606  
 Db 609 HisProGIuHisAlaGlnArgIleGlnSerIleTyrSerArgLeuGlnGlnArgGIyLeu 628  
 QY 1607 CTTAGCAATGCGAGCGGATCCGAGGTGCGAAAGCCAGCCTAGAGATCCAGACAGT 1666  
 Db 629 ArgSerGIuCysGIuLysLeuArgGIyArgLysAlaSerLeuGlnGlnLeuGlnSerVal 648  
 QY 1667 CACTGTGATACACACCCCTGCTATGGGACCAACCCCTCAACCGGAGAGACTAGAC 1726  
 Db 649 HisSerGIuArgHisValLeuLeuTyrGIyThrAsnProLeuSerArgLeuLysLeuAsp 668  
 QY 1727 ACCAAGAGTGTCTGCGCTCCATCAGACCCAGAGATGATGTCTGCTGCTGGGGGC 1786  
 Db 669 AsnGIyLysLeuAlaGlnLeuLeuAlaGlnArgMetPheValMetLeuProCysGIyGly 688  
 QY 1787 ATCGGGGTGACAGTACACACCGCTGTGAATGAGATGACTCTCCAGTCTGTCCGATG 1846  
 Db 689 ValGIyValAspThrAspThrIleTyrAsnGlnLeuHisSerSerAsnAlaAlaArgTyr 708  
 QY 1847 GCAGTGGCTGCTGCTGCTGAGTGGCTTCAGAGTGGCTGCAGAGAGCTCAAGATGGA 1906  
 Db 709 AlaAlaGlySerValThrAspLeuAlaPheLysValAlaSerArgGIuLeuLysAsnGIy 728  
 QY 1907 TTGGCATATCCGGCCCCCAGACACACCGCGAGATCCACAGCCATGGGATTTGTC 1966  
 Db 729 PheAlaValValArgProProGIuHisHisAlaAspHisSerThrAlaMetGIyPheCys 748  
 QY 1967 TTCTCACTGTGTAGCCATCACCCGAAAAGCTCTACAGCAGAGATTTGAAGCTGGGCA 2026  
 Db 749 PhePheAsnSerValAlaIleAlaLysArgGlnLeuGlnGlnSerLysAlaSerLys 768  
 QY 2027 GTTCCTCATCTGAGTGGACATTCACCATGGCAATGAGCAGACAGCGTTCTTCAAT 2086  
 Db 769 IleLeuIleValAspTrpAspValHisHisGlyAsnGIyThrGlnGlnThrPheTyrGln 788  
 QY 2087 GACCCCTGTGCTCTATCTCTCTGATCGTATGATGACAGGGAAGCTTTTCCAGGC 2146  
 Db 789 AspProSerValLeuTyrIleSerLeuHisArgHisAspAspLysAsnPhePheProGIy 808  
 QY 2147 TCTGGGGCTCTGAAGAGTGTGGAGAGCAGCGCTGGGTACATGTGAAGCTGGA 2206  
 Db 809 SerGIyAlaValAspGIuValGIyAlaGlySerGIyGlnGlyPheAsnValAsnValAla 828  
 QY 2207 TGGACAGGAGGTGTGAGCCCCCATTTGAGACGTGAGATACCTTACAGCTTCAGAGACA 2266  
 Db 829 TrpAlaGlyGIyLeuAspProProMetGIyAspProGIuTyrLeuAlaAlaPheArgIle 848

QY 2267 GTGTGATGCCATTTGCCACGAGTTCTCACSTGATGTGCTCTACTGCTCCGGCGGTTT 2326  
 Db 849 ValValMetProIleAlaArgIlePheSerProAlaSerLeuValLeuValSerAlaGlyPhe 868  
 QY 2327 GATGCTGTGAAGACATGTCTCTCTGCTGGGCTACTGTCTGTCACCCACAGATTTT 2386  
 Db 869 AspAlaIleGlnGIyHisProAlaProLeuGlnGIyTyrHisValSerAlaLysCysPhe 888  
 QY 2387 GGGCATGTGACGACGAGTGTGATGACCTGGGAGGGGCGGGGTGGTGGCTGGCGAG 2446  
 Db 889 GlyTyrMetThrGlnGlnLeuMetAsnLeuAlaGlyGIyAlaValValLeuAlaLeuGln 908  
 QY 2447 GAGGCGCATGACTGACCCGCACTGTGATGCTGTGAAGCTTGTCTGCGCTGTCT 2506  
 Db 909 GlyGIyHisAspLeuThrAlaIleCysAspAlaSerGIuAlaCysValAlaIleLeuLeu 928  
 QY 2507 AGTGTAACTGACGCGCTTGGATGAGCAGTCTTGCACAAAGCCCAACATCAACGGA 2566  
 Db 929 GlyAsnArgValAspProLeuSerGIuGlnGIyTyrPlyGlnLysProAsnLeuAsnAla 948  
 QY 2567 GTGGCCAGCCTGAGAAAGTCAATCGAGATCCAGACCAACACTGAGTGTGTCAGAG 2626  
 Db 949 IleArgSerLeuGlnValAlaValIleArgValHisSerLysTyrTrpIlyCysMetGlnArg 968  
 QY 2627 TTGCGCGTGTCTGGCGCGGTCTGTCGAGGGGCCCAAGCAAGTGTGAGACCGAAGCC 2686  
 Db 969 LeuAlaSerCysProAspSerTrpValProArgValProGIyAlaAspLysGIuGlnVal 988  
 QY 2687 GAAT- GTGACGCGCATGGCTTGTCTGTGTGGG-----GCCAAGACGCGCCAAAGCT 2739  
 Db 989 GIuAlaValThrAlaLeuValLeuSerValGIyIleLeuAlaGlnLysP----- 1005  
 QY 2740 GCGGAGCCCGGGAACACAGCCCGGCGGAGAGAGAGCCCATGGAGCGAGAG 2793  
 Db 1006 -----ArgProSerGIuGlnLeuValGIuGln 1015  
 RESULT 10  
 ID AAY97033 standard; Protein; 1428 AA.  
 XX AAY97033;  
 DT 31-OCT-2000 (first entry)  
 XX  
 DE Caspase 8-interacting protein from PAC RPC15-1057120.  
 XX  
 KW Caspase-8 interacting protein; CIP; caspase-8 inhibitor; anti-apoptotic;  
 KW N-acetylglycosamine-6-phosphate deacetylase; p53 TNF-R; neuroprotective;  
 KW tumour necrosis factor receptor; anti-inflammatory; anti-diabetic;  
 KW hepatotropic; virucide; cardiant; anti-ulcer; chromosome 12q31.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200039160-A2.  
 XX  
 PD 06-JUL-2000.  
 XX  
 PE 23-DEC-1999; 99WO-IL00698.  
 XX  
 PR 24-DEC-1998; 98IL-0127721.  
 PR 28-SEP-1999; 99IL-0132105.  
 XX  
 PA (YEDA ) YEDA RES & DEV CO LTD.  
 XX  
 PI Wallach D, Schuchmann M, Goncharov T;  
 XX  
 DR WPI; 2000-452371/39.  
 XX  
 PT New caspase-8 interacting protein capable of interacting with subunit 1  
 PT and/or subunit 2 of caspase-8, useful in the treatment of multiple  
 PT sclerosis with primary oligodendroglipathy, autoimmune uveoretinitis  
 PT and diabetes





```

|||||: ||||||| ||||||| |||||||
Db yAspSerValMetLeuYshISgInCysSerCysGlyAspAsnSerArgHISProGlnH 868
848 yAspSerValMetLeuYshISgInCysSerCysGlyAspAsnSerArgHISProGlnH 868
OY ATGCGGGCGGATCCAGACCATCTGGTCCGGCTGCAGAGACAGCCGCTTAGCAACT 1616
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db ISALeGlyArgGlyLeuInsrtIleTrpSerArgLeuInclunArgGlyLeuArgSerGlnC 888
868 ISALeGlyArgGlyLeuInsrtIleTrpSerArgLeuInclunArgGlyLeuArgSerGlnC 888
OY GCGACGGATCCGAGTCCGAAGCCACGCTAGATAGATCCAGACAGTGCATCTGAAT 1676
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db YSGInCysLeuArgGlyArgGlySAlaserLeuGlnGlnInsrtValHISserGlnA 908
888 YSGInCysLeuArgGlyArgGlySAlaserLeuGlnGlnInsrtValHISserGlnA 908
OY ACCACACCTGCTATGGGACAGTCCCTCAACCGGACGAGAAGCTAGACAGCAAGACT 1736
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db rGHisValLeuLeuYrGlyThrAsnProLeuSerArgLeuYshLeuAspAsnGlySL 928
908 rGHisValLeuLeuYrGlyThrAsnProLeuSerArgLeuYshLeuAspAsnGlySL 928
OY TGCCTGGTCCCAAGCCAGAGATGATGCTGCTGCTGGGCGCATCCGGG----- 1792
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db EALeGlyLeuLeuAlaGlnArgMetPheValMetLeuProCysGlyGlyValProL 948
928 EALeGlyLeuLeuAlaGlnArgMetPheValMetLeuProCysGlyGlyValProL 948
OY ----- 1792
1792 ----- 1792
Db euaIeThrLeuSerAlaPheLeuAlaSerLeuAlaProThrValProGlnGlyLeuSerA 968
948 euaIeThrLeuSerAlaPheLeuAlaSerLeuAlaProThrValProGlnGlyLeuSerA 968
OY ----- 1792
1792 ----- 1792
Db rGValSerTrpGlyLeuYshProProGlyProAsnProYshSerArgProAlaProC 988
968 rGValSerTrpGlyLeuYshProProGlyProAsnProYshSerArgProAlaProC 988
OY ----- 1792
1792 ----- 1792
Db ySPTrpGlyProGlyArgGlyValGlyThrProLeuGlyProGlySerCysValL 1008
988 ySPTrpGlyProGlyArgGlyValGlyThrProLeuGlyProGlySerCysValL 1008
OY ----- 1792
1793 ----- 1793
Db ySPTrpMetMetArgAlaLeuThrLeuAlaProGlnValAspTrpAspTrpIleTrpA 1028
1008 ySPTrpMetMetArgAlaLeuThrLeuAlaProGlnValAspTrpAspTrpIleTrpA 1028
OY ----- 1814
1815 ATGACATGACATCTCCCACTGCTGGCGCATGGAGTGGCGCTGGAGCTGGCT 1874
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db sngIeLeuHISserSerAsnAlaAlaArgTrpAlaAlaGlySerValTrpAspLeuAlaP 1048
1028 sngIeLeuHISserSerAsnAlaAlaArgTrpAlaAlaGlySerValTrpAspLeuAlaP 1048
OY TCAAGGTGGCTGCAGAGAGCTCAAGAAATGATGATTCATCCATCCGCGCCAGAGACCC 1934
1875 TCAAGGTGGCTGCAGAGAGCTCAAGAAATGATGATTCATCCATCCGCGCCAGAGACCC 1934
Db hElyValAlaSerArgGlyLeuYshAsnGlyPheAlaValAlaArgProProGlyHISh 1068
1048 hElyValAlaSerArgGlyLeuYshAsnGlyPheAlaValAlaArgProProGlyHISh 1068
OY ACGCCGAGGAATCCACAGCCATGGATTCGCTTCTTCACTCTGTAGCCATCCGCCAA 1994
1935 ACGCCGAGGAATCCACAGCCATGGATTCGCTTCTTCACTCTGTAGCCATCCGCCAA 1994
Db ISAlaAspHISserThrAlaMetGlyPheCysPhePheAsnSerValAlaIleAlaCysA 1088
1068 ISAlaAspHISserThrAlaMetGlyPheCysPhePheAsnSerValAlaIleAlaCysA 1088
OY AACTCTACAGACAGAAGTTGAACGTGGCAAGTCTCTATCTGGACTGGGACATTCACC 2054
1995 AACTCTACAGACAGAAGTTGAACGTGGCAAGTCTCTATCTGGACTGGGACATTCACC 2054
Db rGlnLeuGlnGlnInsrtLysAlaSerLysIleLeuIleValAspTrpAspValHISh 1108
1088 rGlnLeuGlnGlnInsrtLysAlaSerLysIleLeuIleValAspTrpAspValHISh 1108
OY ATGGCAATGGCCAGCCAGCGGCTTCTACAAATGACCCCTGCGCTCATCTCTGCTC 2114
2055 ATGGCAATGGCCAGCCAGCGGCTTCTACAAATGACCCCTGCGCTCATCTCTGCTC 2114
Db ISGlyAsnGlyThrGlnGlnInsrtPheYrGlyAspProSerValLeuYrTlleSerLeuH 1128
1108 ISGlyAsnGlyThrGlnGlnInsrtPheYrGlyAspProSerValLeuYrTlleSerLeuH 1128
OY ATCGCTATGACACAGCGGAATCTTCTCCAGGCTCTGGGCTCTGAGAGAGTGTGTGAG 2174
2115 ATCGCTATGACACAGCGGAATCTTCTCCAGGCTCTGGGCTCTGAGAGAGTGTGTGAG 2174
Db ISArgHISasprspGlyAsnPhePheProGlySerGlyAlaValAspGlyValGlyAla 1148
1128 ISArgHISasprspGlyAsnPhePheProGlySerGlyAlaValAspGlyValGlyAla 1148
OY GACCGAGCGCTGGGGTACATGTGAACGTGGCATGGACAGAGAGTGGAGCCCCCATGG 2234
2175 GACCGAGCGCTGGGGTACATGTGAACGTGGCATGGACAGAGAGTGGAGCCCCCATGG 2234
Db lYSerGlyGlnGlyPheAsnValAsnValAlaIleTrpAlaGlyLeuAspProPheMeG 1168
1148 lYSerGlyGlnGlyPheAsnValAsnValAlaIleTrpAlaGlyLeuAspProPheMeG 1168
OY GAGAGGTGAGTACCTTACAGCTTCAGAGACAGTGGTGAATGGCCATGGCCAGAGTTCT 2294
2235 GAGAGGTGAGTACCTTACAGCTTCAGAGACAGTGGTGAATGGCCATGGCCAGAGTTCT 2294
Db lYAspProGlyTrpLeuAlaAlaPheArgIleValAlaMetProIleAlaArgGlyPheS 1188
1168 lYAspProGlyTrpLeuAlaAlaPheArgIleValAlaMetProIleAlaArgGlyPheS 1188
OY CACGTGATGTGCTAGTCTCCGCGGCTTGAATGCTGTGAAGACATCTGTCTCTC 2354
2295 CACGTGATGTGCTAGTCTCCGCGGCTTGAATGCTGTGAAGACATCTGTCTCTC 2354
Db eSProAspLeuValLeuValSerAlaGlyPheAspAlaAlaGlnGlyHISProAlaProL 1208
1188 eSProAspLeuValLeuValSerAlaGlyPheAspAlaAlaGlnGlyHISProAlaProL 1208
OY TGGGTGGCTACTCTGTACACCGCCAGATGTTTGGCCACTTGACACAGGACGTGATGACC 2414
2355 TGGGTGGCTACTCTGTACACCGCCAGATGTTTGGCCACTTGACACAGGACGTGATGACC 2414

```

```

Db 1208 euglyGlyTrpHisValSerAlaLysCysPheGlyTyTrpMetThrGlnGlnLeuMetAsnL 1228
OY 2415 TGCAGAGGGCGCGGTGCTGTGGCTTCGAGAGGAGCCATGATCTGACCGCATCTGTG 2474
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1228 euaIeGlyLysAlaValIleAlaLeuLeuGlnGlyLysHisAspLeuThrAlaIleCysA 1248
OY 2475 ATGCTCTGAGACCTGTGTCTCGGCTCTGCTCAAGTGAAGCTGACGCCCTTGATGAGG 2534
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1248 sPAlaserGlnAlaCysValAlaIleAlaLeuLeuGlnYshnArgValAspProLeuSerGlnG 1268
OY 2535 CAGTCTGACGAGAAAGCCCAACATCAAGCAGTGGCCACCGCTAGAGAAATCATCGA 2594
1268 lueGlyTrpYshGlnYshProAsnLeuAsnAlaIleArgSerLeuGlnAlaValIleArgY 1288
OY 2595 TCCAGAGCAACACATGAGCTGTGTGTCAGAACTTCGCCCTGCTGTGGCGGCTCCCTGC 2654
1288 aIHISerLys----- 1291
Db 2655 GAGGGGCCCAAGCAGTGTGAGACCGAAGCCGAAATGTGAAGCCCATGCGCTTGTCTGT 2714
1292 ----- -CysGlyAspGlyThrLeuAlaGlnLeu----- 1300
OY 2715 GGTGGGGCCGGAACAGCCCAAGCTGCGGCAGCCCGGGAACACAGCCCGGACAG 2774
Db 1301 ----- 1314
OY 2775 GGAGCCCATGAGAGCAGACCTGCCCCCTGTGAC 2806
Db 1314 lInleuGlnYpHeArgCysGlnProGlyAsp 1324

RESULT 11
AAU99662
ID AAU99662 standard; Protein; 855 AA.
AC
XX AAU99662;
AC
XX
DT 07-OCT-2002 (first entry)
XX
DE Human Histone deacetylase Isoform 7.
XX
KW Human; histone deacetylase; HDAC-7; cancer; cytostatic;
KW antisense; tumour suppressor; cell proliferation; tumour;
KW programmed cell death; necrotic cell death.
XX
OS Homo sapiens.
XX
FH
FT Key Location/Qualifiers
FT Misc-difference 54 /note= "Encoded by ACG"
FT Misc-difference 117 /note= "Encoded by CCG"
FT Misc-difference 120 /note= "Encoded by CCG"
FT Misc-difference 134..135 /note= "Encoded by CTG"
FT Misc-difference 137 /note= "Encoded by CTTCCG"
FT Misc-difference 140 /note= "Encoded by GCT"
FT Misc-difference 148 /note= "Encoded by CCG"
FT Misc-difference 163 /note= "Encoded by TTC"
FT Misc-difference 169 /note= "Encoded by CAT"
FT Misc-difference 186 /note= "Encoded by CCG"
FT Misc-difference 211 /note= "Encoded by CCC"
FT Misc-difference 261 /note= "Encoded by CCC"
FT Misc-difference 270 /note= "Encoded by ACC"
FT Misc-difference 270 /note= "Encoded by CCT"

```

FT Misc-difference 298 /note- "Encoded by TTA"  
 FT Misc-difference 384 /note- "Encoded by CCG"  
 FT Misc-difference 409..410 /note- "Encoded by TTGCTC"  
 FT Misc-difference 428 /note- "Encoded by ACT"  
 FT Misc-difference 430 /note- "Encoded by CTG"  
 FT Misc-difference 434 /note- "Encoded by CTG"  
 FT Misc-difference 479 /note- "Encoded by CTG"  
 FT Misc-difference 520 /note- "Encoded by ACC"  
 FT Misc-difference 534 /note- "Encoded by CCG"  
 FT Misc-difference 567..568 /note- "Encoded by CTG"  
 FT Misc-difference 595 /note- "Encoded by CTCCTG"  
 FT Misc-difference 746 /note- "Encoded by NCC"  
 FT Misc-difference 773 /note- "Encoded by ACG"  
 FT Misc-difference 845 /note- "Encoded by CNG"  
 FT Misc-difference 845 /note- "Encoded by TCG"  
 PN US2002061860-A1.  
 XX 23-MAY-2002.  
 XX 06-AUG-2001; 2001US-0817913.  
 PF 24-MAR-2000; 2000US-192157P.  
 XX (L12Z/) LI Z.  
 PA (BONE/) BONFILS C.  
 PA (BEST/) BESTERMAN J.  
 XX LI Z, Bonfils C, Besterman J;  
 XX WPI: 2002-507650/54.  
 DR N-PDB; ABK87721.  
 PT Agent that specifically inhibits an isoform of histone deacetylase,  
 PT useful for treating cancer and other cell proliferative diseases,  
 PT preferably comprising an antisense oligonucleotide  
 XX Disclosure; Fig 7A: 60pp; English.  
 PS The invention relates to an agent that inhibits an isoform of histone  
 CC deacetylase (HDAC-1 to HDAC-8) but not all isoforms, e.g. an  
 CC antisense oligonucleotide. Also included are inhibiting an HDAC isoform  
 CC in a cell by treatment with the agent, identifying an HDAC isoform that  
 CC is required for induction of cell proliferation or differentiation and  
 CC inhibiting cell proliferation by treatment with two antisense  
 CC oligonucleotides or small molecules that inhibit a specific HDAC  
 CC isoform, or antisense oligonucleotide or small molecules that inhibit  
 CC DNA methyltransferase. The agent therefore acts as a tumour suppressor.  
 CC The agents are used to treat diseases of cell proliferation and  
 CC differentiation (e.g. cancer and tumours), by inducing growth retardation,  
 CC growth arrest or programmed/necrotic cell death, specifically neoplastic  
 CC cell proliferation in humans. The agents are selective for particular  
 CC isoforms, compared to known inhibitors which are not selective.  
 CC The present sequence represents the HDAC-7 isoform.  
 XX  
 SQ Sequence 855 AA;

Alignment Scores: 2.33e-104 Length: 855  
 Pred. No.: 1671.50 Matches: 384  
 Score:

Percent Similarity: 53.47% Conservative: 101  
 Best Local Similarity: 42.34% Mismatches: 195  
 Query Match: 31.27% Indels: 227  
 DB: 23 Gaps: 22  
 US-09-502-945-2 (1-2885) x AAU99662 (1-855)  
 QY 11 TTGTGAGTCAAGTCAAGAGCCACGAGCGGCGCTCAACCAATTCCTCCACAGACACCC 70  
 DB 101 LeuGluArgThrValHisProAsnSerProGlyLeuProGlyArgThrLeuGluProIle 120  
 QY 71 AAATGCTGGGAGCCACACATGCTTTTGGACAGAGTTCCTCCACAGAGCGGCC 130  
 DB 121 GluThrGluGlyAlaThrArgSerMetLeuSerSerPheLeuProValProSerIle 140  
 QY 131 CTTGGAGCGCTCCCTCTACAACTGCTTTGGCTGGCCCTACGACAGTCCAGAGAC 190  
 DB 141 ProSerAspProProGluHis----- 147  
 QY 191 TTCCCTCCGCAAAACAGCCTGTGAACCACTTGAAGGCGTTCAGAGCTAAACAG 250  
 DB 148 PheProLeuArgGlySerThrValSerGluProAsnLeuGlySerGlyProLys--- 166  
 QY 251 AAGGTGCTGAGCGAGACAGAGTCCCTGCTGCGACAGATGGAGCTTTATTAGC 310  
 DB 167 LysSerLeuGluArgArgLysAsnProLeuLeuArgLysGluSerAlaPro-----Pro 184  
 QY 311 ACCTTTAAGAAAGAGAGCTGTGAGATCATCAGAGTCCCGGCGCTGCTGCTGT 370  
 DB 185 SerLeuArgArgArgProAlaGluThrLeuGlyAspSerSerProSerSer----- 202  
 QY 371 AACAGCGCACCGGCTCCGCGCCAGCTTCCCAACAGTCCACAGACACCATGCTGAG 430  
 DB 203 ---SerThrProAlaSerGlyCysSerSerProAsnAspSer-----Glu 216  
 QY 431 AATGCTTTACTGCTGAGTCCACATCCCACTGAGATGCTCCTCAGACCGAGCC 490  
 DB 217 HisGly----- 218  
 QY 491 CTCCTCTGAGACAGTCCCAACCACTTACAGCTTACAGTCTCTCTGCTCCCAAC 550  
 DB 219 ----- 220  
 QY 551 ATCTCCCTAGGAGCTGAGCGCCAGCGTCACTGTACCAACTCACACTCAGTCCCTCCCG 610  
 DB 221 ProIleLeuGly----- 224  
 QY 611 AAGCTGTGACACAGACAGAGCGGAGCGAGCGCCCTCAGTCCCTGCGGAGGTCGC 670  
 DB 225 -----AspSerAspArgArgThrHisProThrLeuGlyProArgGly 238  
 QY 671 ACGCTGACCGGCAAGTTCATGAGCAGCATCTCTATTCT----- 709  
 DB 239 ProIleLeuGlySerProHisThrProLeuPheLeuProHisGlyLeuGluProGluAla 258  
 QY 710 ---GGTGGCTGCGGGCGGTGGACCTGAGGGCGACGGAGCGGCGGATGCTCC 766  
 DB 259 GlyLeuGlySer-----ProSer 264  
 QY 767 CTGCTGACAGATGCTGTCTGTGAGCAGCGCCGCGAGACAGACACCTCATTTGCTGTG 826  
 DB 265 ArgLeuGluProIleLeuLeuAspProSerGlySerHisAlaProLeuLeuThrVal 284  
 QY 827 CCACTTCAGCGGAGCTCCCACTGATGAGGGGTGAACSTGTGGCCACAGATGCGGACG 886  
 DB 285 ProGlyLeuGlyProLeuProHisPheAlaGluSerIleMetThrGluArgLeu 304  
 QY 887 GTAGGCAAGCTCCGCGGATCGGCGCTGAGCGGCGGACGAGTCCAGTCCGCGGCGAG 946  
 DB 305 SerGlySer---GlyLeuHisThrProLeuSerArgThrArgSerGluProLeuProPro 323  
 QY 947 AGTCCCGGAGCC-----CTGAGCAGCTGCTGTCATG 976  
 DB 111 111 111



PI Oka T., Sogai T., Nishikawa T., Hayashi K., Saito K., Yamamoto J;  
PI Ishii S., Sugiyama T., Wakamatsu A., Nagai K., Otsuki T;  
XX WPI; 2001-318749/34.  
XX  
PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and/or the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
PS  
PS Claim 8; SEQ ID 11098; 2537bp + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.



[illegible]

Db	727	uLeuLeuInlInlInlInlUGU-----GluLeuAlaLysSerLeuMetAlaArg	742
QY	1303	TGGTGTCTGGATACAAAAAGTGTCTCAGATAGCCCAACCGCTGCACCTTTGACGTGA	1362
Db	742	oLeuSer-----ArgThrLeuSerSerProLeuValProLeuGlyProHisGlyLeuSe	760
QY	1363	CCAAACCGGCC-----1372	
Db	760	rGlnIleProAspThrGlyGlnGlnProAlaProIleAlaThrSerSerSerAlaAspHi	780
QY	1373	-----CTACAGCTGGCCACTGTGGTCCCAACAGCCCTGGCCGTAC--	1414
Db	780	sILleProValAlaSnLeuSerLeuProHisArgIleHisArgGlnLeuMetSerThrIle	800
QY	1415	-----CAATCTCCCGCTGGTGGCCCTGGGGCGGATAAACAACCCCC	1455
Db	800	uTrpAlaSerGlnLeuAlaArgAsnHisGlnProSerAlaLysArgI-----SerIlePr	817
QY	1456	AGACCAACCCCTCAAGACACTCTTTCACCAACAAAGTGGTGTCTGCAGACACTGCATCTAA	1515
Db	817	o-----HisLysValThrThrGlnGlyLeuAlaIaLysAspProIleuMetLeuLys	832
QY	1516	GCACCACTGCATGTGGCGGAACACACACAGCTGCACCTGCAGACTGTGGCCGATCCAGAG	1575
Db	832	shISerCysLileCysGlyAspAsnAlaGlnHisProGlnHisSerGlyArgLeuGlnSe	852
QY	1576	CATCTGGTCCCGCGGTGCAGAGACAGCCGCTGTACCAAGTGGCGAGCGGATCCGAGTGC	1635
Db	852	rValTrpAlaArgLeuLysnGlnuTrpAspLysValLysArgCysAspArgLysLeuAlaArg	872
QY	1636	CAAAAGCCAGCTAGATAGATATCCAGACAGTGCACCTGTGAATCCACACCTCTCTATAG	1695
Db	872	gLyAlaThrGlnGlnGlnLeuGlnThrValHisThrGlnAlaHisAlaMetLeuPheGln	892
QY	1696	GACCACTGCCCTCAACCGCGCAAGAACTAGACACAAAGAAAGTTGTCTGGTCCATAGCCA	1755
Db	892	ySer-----AsnGlnCysGlnLeuSerAlaArgProLysLeuGlnAsnThrLeuSerAl	909
QY	1756	GAAGATGTACTGTGTGCGCTGTGGCGGGGATCGGGGGAGACAGTACACCTGTGGAA	1815
Db	909	Aser---PheValArgLeuSerCysGlyGlyLeuGlyValAspLysLeuAspThrThrPAs	928
QY	1816	TGACATGCACCTCTCCAGTGTGTGTGCACATGCAGTGGGGCTGTCTGGAGCTGGCCTT	1875
Db	928	nGlnHisHisThrAlaThrAlaAlaAlaArgMetAlaAlaGlyCysValIleAspLeuAlaLe	948
QY	1876	CAAGGTCTGTGCAGAGAGACTCAAGAAATGGATTGTGCATATCATCCGGCCCAAGACACA	1935
Db	948	uLysThrAlaLysGlyAspLeuAlaArgAsnGlyPheAlaValAlaArgProGlyHisHi	968
QY	1936	CGCGGAGGAATCCACACCAATGGGATGTCTCTTCACCTGTGATCCATCCACCGCAAA	1995
Db	968	sAlaGlnAlaAsnLeuAlaMetGlyPheCysPhePheAsnSerIleAlaAlaIleAlaLys	988
QY	1996	ACTCTACAGCAAGAAGTTG---AAGCGGCAAGAGTCCATATCGTGGACTGGGACATTTCA	2052
Db	988	sLeuLeuAlaArgIleArgMetProGlnuAlaArgAlaIleLeuIleValAspTrpAspValHi	1008
QY	2053	CCATGGCAATGGCAACCAAGAGCGTTTTCACAATGACCCCTGTGCTCTACATCTCTCT	2112
Db	1008	shISerLysnGlyThrGlnGlnAlaPheArgGlnSerProAspIleLeuThrLysSerTi	1028
QY	2113	GCATCGCATATCAACAAGGAAATCTTTCACAGCTGTGGGGGCTCGCAAGACATGTGGGG	2172
Db	1028	eHisArgIleAspAspLysnAspPhePheProGlyTrnGlyGlyProThrGlnCysGlySe	1048
QY	2173	AGGACCAAGCGCTGGGTATACATGTGAACGTGGCATGGAGAGAGTGGAGACCCCCCAT	2232
Db	1048	rGlyAlaGlyLeuLysPheAsnValAsnIleSerTrpSerGlyAlaLeuAsnProIole	1068
QY	2233	TGGAGACGTGGATACCTTACAGCTTTCAGACAGATGTGATGCCCATTTGCCACGAGATT	2292

Db 1068 UGLYSPALAGLUTYRILEALALAPHEARGLTHRVAlValMetProILeAlARSerph 1088  
 QY 2293 CTCACCTGATGTGTCTAGTCTCCGCCGGTGTGATGCTGTGAAGACATGTGTCTCC 2352  
 Db 1088 easnProasprileValLeuValSerSerCylpheasprAlaAlaIlnGlnHisProAlaIrr 1108  
 QY 2353 TCTGGGTGGCTACTGTGCACCGCCAGATGTTTGGCCACTTGGACCCAGCCAGCTGATAC 2412  
 Db 1108 OleuAllylGlyTRHValSerProAlaCysPheGlyPheMetTrpArgGluLeuLeu 1128  
 QY 2413 CTTGACGAGGGGGCGGGTGTGCTGGCCCTGAGGAGGACCATGACTTACCCAGCCATCTG 2472  
 Db 1128 nleuAlasnGlyValValLeuAlaLeuGlnGlyGlyTRAPRleAlaAlaIleCyl 1148  
 QY 2473 TGATGCTCTGTAAGCTTGTGTCTGCGCTCTGCTCACTGTAAGCTGACCCCTTGATGA 2532  
 Db 1148 sasprerAlaGlnGlnCysValArgAlaLeuLeuGlnAspProAlaAlaProILeAla 1168  
 QY 2533 GGCAGTCTTGGCAGCAAAAGCCCAATCATCAGCAGTGGCCAGCTAGAGAAAGTATCGA 2592  
 Db 1168 sAlaGluLeuGlnArgProProCysGlnAsnAlaIleAsnTrpLeuGlnGlyThrIleAl 1188  
 QY 2593 GATCCAGAGCAACACTGAGCTGTGTGCGAAGTTTCGCCGCTGGTGGCGCCGCTCC 2652  
 Db 1188 AlleGlnGlnTrpHisTrpProCysValArgMetLeuGlnHisTrpValGlyLeuSerAl 1208  
 QY 2653 GCGAGGGGCCCCAAGCAGGTGAGACCCAGAGAACCCGAA 2689  
 Db 1208 aleuGlnThrLeuLysValGlnHisAspGlnSerGlu 1220

## RESULT 14

AA97032 ID AAY97032 standard; Protein: 574 AA.

AC AAY97032;

DT 31-OCT-2000 (first entry)

DE Caspase-8-interacting protein clone p74.

KM Caspase-8 interacting protein; CIP; caspase-8 inhibitor; anti-apoptotic;

KM N-acetylglucosamine-6-phosphate deacetylase; p55 TNF-R; neuroprotective;

KM tumour necrosis factor receptor; anti-inflammatory; anti-diabetic;

KM hepatotropic; virucide; cardiant; anti-ulcer; chromosome 12q31.

OS Homo sapiens.

PN WO200039160-A2.

PD 06-JUL-2000.

PE 23-DEC-1999; 99WO-IL00698.

PR 24-DEC-1998; 98IL-0127721.

PR 28-SEP-1999; 99IL-0132105.

PA (YEDA ) YEDA RES & DEV CO LTD.

PI Wallach D, Schuchman M, Goncharov T;

DR WPI; 2000-452371/39.

DR N-PSDB; AAA51856.

XX New caspase-8 interacting protein capable of interacting with subunit 1

PT and/or subunit 2 of caspase-8, useful in the treatment of multiple

PT sclerosis with primary oligodendroglipathy, autoimmune uveoretinitis

PT and diabetes

XX Claim 3; Fig 5B; 125pp; English.

XX A caspase-8 interacting protein (CIP), capable of interacting with

CC subunit 1 and/or 2 of caspase-8, was identified by a yeast two hybrid

CC screen in human placental cDNA clone P74. A similar sequence was also

CC found in a genomic clone (RPC15-1057120) which localizes to human  
 CC chromosome 12q31. clone p74 shows homology to a highly homologous region  
 CC of mouse and human histone deacetylases. CIPs inhibit caspase-8 and human  
 CC p55 tumour necrosis factor receptor (TNF-R) induced apoptosis in HKK  
 CC 293-T and HeLa cells. The CIP, CIP-targeted ribozymes, antisense  
 CC oligonucleotides or antibodies are useful for modulating caspase-8  
 CC activity, TNF-R or Fas-mediated effects, or apoptosis. They are also  
 CC useful in the treatment of multiple sclerosis with primary  
 CC oligodendroglipathy, autoimmune uveoretinitis, diabetes, lupus,  
 CC autoimmune myocarditis I, hepatitis C virus (HCV) mediated chronic  
 CC hepatitis, chronic gastritis (e.g. type A gastritis), mixed connective  
 CC tissue disease (MCTD), Crohn's disease, or ulcerative colitis. CIP are  
 CC also useful for the isolation, identification and cloning of other  
 CC proteins of the same class.

CC SQ Sequence 574 AA;

CC Alignment Scores:

CC Pred. No.: 9,14e-98 Length: 574

CC Score: 1573.00 Matches: 317

CC Percent Similarity: 68.65% Conservative: 75

CC Best Local Similarity: 55.52% Mismatches: 109

CC Query Match: 29.42% Indels: 71

CC DB: Gaps: 9

US-09-502-945-2 (1-2885) x AAY97032 (1-574)

QY 1202 GAGGACGAGGAAGAGGAGGAGGAGAGATGATCAGATTAAAGACGAGGAG 1261

Db 29 GlnAspLeuGlnThrAspGly----- 35

QY 1262 GCGGAGAGTGT-----GCTGAGGAGGGGCCGACTTGAGAGACCTGTGATAC 1315

Db 36 GlyGlyProGlyGlnValValAspAspGlyLeuGlnHisArgGluLeuGlnHisGlyGln 55

QY 1316 AAAAACTGTCTGAGATGCCCAACCGCTGCACTTGTGACGTACCAGCGCC--- 1372

Db 56 -----ProGluAlaArgGlyProAlaProLeuGln-----GlnHisProGln 69

QY 1372 ----- 1372

Db 70 ValLeuLeuTrpGlnGlnArgLeuAlaGlyArgLeuProAlaGlySerThrGlyAsp 89

QY 1373 -----CTGACCTGGCCACTGTGTCCCAAGCCCTGGGCCCTGATCCTCC 1423

Db 90 ThrValLeuLeuProLeuAlaGlnGlyGlnHisArgProLeuSerArgAlaGlnSerSer 109

QY 1424 CTTGTGCCCCCTGGGGGCGATGAAGACCC----- 1453

Db 110 ProAlaAlaProAlaSerLeuSerAlaProGluProAlaSerGlnAlaArgValLeuSer 129

QY 1454 ---CCAGACCAACCCGCTGATGAGTCTGACAGACAGGACCTGATACACACCTTC 1507

Db 130 SerSerGluTrpProAlaArgThrLeuProRheTrpGlnGlyLeuIleTyrAspSerVal 149

QY 1508 ATGCTAAAGACACACAGTGCATGTGGGGAACACACACAGTGCACCTGAGACATGTGGCCG 1567

Db 150 MetLeuLysHisGlnCysSerCysGlyAspAsnSerArgHisProGlnHisAlaGlyArg 169

QY 1568 ATCCAGACATGTGTCCTGGCTGCAGAGACAGAGCTGTGATACAGTGGAGCGATC 1627

Db 170 IleGlnSerIleTyrSerArgLeuGlnGlnArgGlyLeuArgSerGlnCysGlyLeu 189

QY 1628 CGAGGTGCAGAAAGCCGCTGATGAGTCTGACAGACAGGACCTGATACACACCTTC 1687

Db 190 ArgGlyArgGlyAlaSerLeuGlnGlnLeuGlnSerValHisSerGlnAlaGlnHisValLeu 209

QY 1688 CTTATGGAGACAGTCCCTCAACCGGACAGAGTGTAGACAGAAAGTGTGCTGCTCC 1747

Db 210 LeuTyrGlyThrAsnProLeuSerArgLeuLysLeuAspAsnGlyLysLeuAlaGlyLeu 229

QY 1748 ATCAGCCAGAAAGTATGCTGTGCTGCTGGGGCATCGGGGTGGACAGTACACC 1807



```

Db 230 LeuAlaGlnArgMetPheValMetLeuProCysGlyValGlyValAlaSerThrAspThr 249
QY 1808 GTGGGAAATGATGACCTCTCCAGTGTCTGTGGCATGGCAGTGGGCTGCTCTGGAG 1867
Db 250 ILerPrasnGlnLeuHisSerSerAsnAlaAlaArgTrpAlaAlaGlySerValThrAsp 269
QY 1868 CTGGCCCTTCAGAGTGGCTGGAGAGACCTCAAGAAATGGATTTGGCCATCCGCCGCCCA 1927
Db 270 LeuAlaPheLysValAlaSerArgGlnLeuLysAsnGlyPheAlaValAlaArgProPro 289
QY 1928 GGAACACACCCGAGAGAAATCCACAGCCATGGAGATTGCTTCTTCAACTGTGATGCCATC 1987
Db 290 GLyHisHisAlaAspHisSerThrAlaMetGlyPheCysPhePheAsnSerValAlaIle 309
QY 1988 ACCGCAAAACTCTACAGCAAGAACTTGAAGCTGGGCAAGCTCCATCTGTGATGGGAC 2047
Db 310 AlaCysArgGlnLeuGlnGlnGlnSerLysAlaSerLysIleLeuIleValAlaSerPasp 329
QY 2048 ATTACACCATGGCAATGGCAGCCAGCAGCGCTTACAAATGACCCCTGTGCTTACATC 2107
Db 330 ValHisHisGlyAsnGlyThrGlnGlnGlnThrPheTyrGlnAspProSerValLeuTyrIle 349
QY 2108 TCTCTGCATGCTATGACAAAGGGAACTTCTTCCAGGCTCTGGGCTCTCTGAAGAGTT 2167
Db 350 SerLeuHisArgHisAspArgLysAsnPhePheProGlySerGlyAlaValAlaSerGluVal 369
QY 2168 GGTGGAGGACACGGCGTGGGGTACATGTGAAGCTGGCATGGACAGAGAGTGTGACCC 2227
Db 370 GLyAlaGlySerGlyGlyGlyLysPheAsnValAlaSerAlaGlyGlyLeuAspPro 389
QY 2228 CCCATTTGAGACAGTGTGAGTACCTTACAGACCTTACAGACAGTGTGATGGCAGCCAC 2287
Db 390 PrometCylsSprProGluTyrLeuAlaAlaPheArgIleValAlaMetProIleAlaArg 409
QY 2288 GAGTTTCACTGATGTGTGCTAGTCTCCGCGGGTTTGAATGCTGTGAAGACATCTG 2347
Db 410 GluPheSerProAspLeuValLeuValSerAlaGlyPheAspAlaGlyGlyLysIlePro 429
QY 2348 TTTCTCTGGTGGTCTACTGTCTGACCGCCAGATGTTTGGCCATGGACAGGCACTG 2407
Db 430 AlaProLeuGlyGlyTyrHisValSerAlaLysCysPheGlyTyrMetThrGlnGlnLeu 449
QY 2408 ATGACCCCTGGCAGGGGCGCGGTGCTGCTGGAGGGAGGAGCCATGATTCAGCCGCC 2467
Db 450 MetAsnLeuAlaGlyGlyAlaValAlaLeuAlaLeuGlnGlyGlnHisAspLeuThrAla 469
QY 2468 ATCTGTATGCTCTGAAGCTGTGTCTCGGCTCTGCTCACTGTAAGCTGACCCCTTG 2527
Db 470 IleCysAspAlaSerGlnAlaCysValAlaAlaLeuLeuGlnLysAsnArgValAlaSerProLeu 489
QY 2528 GATGAGGACAGTCTTGCAGCAAAAGCCCAATCAACGACAGTGGCCAGCTAGAGAAAGTC 2587
Db 490 SerGlnGlnGlyTyrPylsGlnLysProAsnLeuAsnSerIleArgSerLeuGlnAlaVal 509
QY 2588 ATCGAGATTCAGAGCAAAACCTGGAGCTGTGTGCAGAAAGTTCCGCGCTGTCTGGCGCG 2647
Db 510 IleArgValHisSerLysTyrTrpGlyCysMetGlnArgLeuAlaSerCysProAspSer 529
QY 2648 TCCCTGGAGGGGCGCCAGAGGAGTGAAGAGCAAGCCCAAT-GTGAAGCCATGGCC 2706
Db 530 TrpValProArgValProGlyAlaAspLysGlnValGlnAlaValAlaThrAlaLeuAla 549
QY 2707 TTGTGTGTGGTGGG-----GCCGACAGGGCCCAACCTGGGCGAGCCGGGAAACACAGC 2760
Db 550 SerLeuSerValGlyIleLeuAlaGlnAsp----- 559
QY 2761 CCCAGCGCGCAGAGAGCCATGAGCAGAGAG 2793
Db 560 ---ArgProSerGlnGlnLeuValGlnGlnGln 569

```

RESULT 15  
 AAB42845  
 ID AAB42845 standard; Protein; 541 AA.

```

XX AC AAB42845;
XX DT 08-FEB-2001 (first entry)
XX DE Human ORFX ORF2609 polypeptide sequence SEQ ID NO:5218.
XX KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerable; antipruritic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antineumatic; antihypoid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX KW Homo sapiens.
XX OS WO200058473-A2.
XX PN 05-OCT-2000.
XX PD 31-MAR-2000; 2000WO-0508621.
XX PE 31-MAR-1999; 99US-0127607.
XX PR 02-APR-1999; 99US-0127636.
XX PR 05-APR-1999; 99US-0127728.
XX PR 30-MAR-2000; 2000US-0540763.
XX PA (CURA-) CURAGEN CORP.
XX PI Shinkets RA, Leach M;
XX DR WPI, 2000-602362/57.
XX DR N-PSDB; AAC77034.
XX PT Novel nucleic acids and peptides derived from open reading frame X,
XX PT useful for treating e.g. cancers, proliferative disorders,
XX PT neurodegenerative disorders and cardiovascular disease -
XX PS Claim 11; Page 4402-4403; 5507pp; English.
XX CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
XX CC antipruritic; antiparkinsonian; nootropic; neuroprotective;
XX CC osteopathic; anticonvulsant; thrombolytic; coagulant; vasotropic;
XX CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX CC antidiabetic; hypotensive; dermatological; immunosuppressive;
XX CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
XX CC antihypoid; and antianemic. The sequences can be used for determining
XX CC the presence of or predisposition to, or preventing or treating
XX CC pathological conditions associated with an ORFX-associated disorder. The
XX CC nucleic acids can be used to express ORFX proteins in gene therapy
XX CC vectors. The proteins and nucleic acids may be used to treat cancers,
XX CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX CC graft vs host disease, cardiovascular disease, diabetes mellitus,
XX CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
XX CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX SQ Sequence 541 AA;

```

Alignment Scores:



GenCore version 5.1.4 p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: March 21, 2003, 12:20:42 ; Search time 53.2241 Seconds

(without alignments)  
12016.549 Million cell updates/sec

Title: US-09-502-945-1

Perfect score: 2732  
Sequence: 1 cttctgcatcgcacgcagaaa.....aaatgacattttaagaaga 1552

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+ n2p.model -DEV=xlp  
-O/cgn2.1/USFTO.spool/US0502945/runat\_14032003\_101058\_19113/app.query.fast\_1.10979  
-DB=SPREMBL\_21 -QFMT=fastan -SUFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR.SCORE=pcr -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pcr -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USRR=US0502945 @cgn.1.1.565 @runat.14032003.101058.19113 -NCPU=6 -ICPU=3  
-NO\_XLPXY -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -MAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: SPREMBL\_21:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mammal:\*  
8: sp\_mhc:\*  
9: sp\_organelle:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_virus:\*  
14: sp Vertebrate:\*  
15: sp\_unclassified:\*  
16: sp\_virus:\*  
17: sp\_bacteriophage:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2062	75.5	414	4	060527 homo sapien

2	935	34.2	468	11	09CSM9	09CSM9 mus musculu
3	841	30.8	393	4	09POF1	09POF1 homo sapien
4	521	19.1	379	11	08R3K0	08R3K0 mus musculu
5	343	12.6	314	11	09CYH0	09CYH0 mus musculu
6	272.5	10.0	178	5	076329	076329 dictyostell
7	256.5	9.4	1790	3	007380	007380 saccharomyc
8	254	9.3	2017	5	094992	094992 drosophila
9	254	9.3	2056	5	09M0W8	09M0W8 drosophila
10	254	9.3	2057	5	094987	094987 drosophila
11	253.5	9.3	1931	13	091973	091973 coturnix co
12	253.5	9.3	3259	4	014789	014789 homo sapien
13	252	9.2	1022	11	09WVC0	09WVC0 mus musculu
14	251	9.2	876	5	09BMW8	09BMW8 sarcoscyt s
15	250.5	9.2	891	13	09M0H5	09M0H5 gallus gall
16	250.5	9.2	1931	13	0910C5	0910C5 gallus gall
17	250.5	9.2	3616	13	09W6V0	09W6V0 gallus gall
18	249.5	9.1	1207	13	002015	002015 gallus gall
19	249	9.1	1294	5	09VXU1	09VXU1 drosophila
20	249	9.1	1398	5	0960D0	0960D0 drosophila
21	248	9.1	2611	11	091208	091208 mus musculu
22	246.5	9.0	682	5	096720	096720 dermatophag
23	246.5	9.0	1581	4	092614	092614 homo sapien
24	246.5	9.0	1992	13	004834	004834 xenopus lae
25	245	9.0	1583	4	015045	015045 homo sapien
26	244.5	8.9	1964	5	08SWQ7	08SWQ7 loligo peal
27	244	8.9	688	4	096NL6	096NL6 homo sapien
28	244	8.9	764	13	091411	091411 gallus gall
29	243.5	8.9	625	6	002717	002717 bos taurus
30	243.5	8.9	1229	5	09NJ22	09NJ22 aequipecten
31	243.5	8.9	1243	5	09NJ21	09NJ21 aequipecten
32	243.5	8.9	1253	5	09NJ20	09NJ20 aequipecten
33	243.5	8.9	1588	11	09ESK9	09ESK9 mus musculu
34	243.5	8.9	1951	5	017042	017042 aequipecten
35	241.5	8.8	1939	13	09PTP2	09PTP2 gallus gall
36	241.5	8.8	2442	4	060588	060588 homo sapien
37	241.5	8.8	2442	4	09H4S0	09H4S0 homo sapien
38	241	8.8	1313	4	075033	075033 homo sapien
39	241	8.8	1927	5	025142	025142 halocynthia
40	241	8.8	1940	5	002456	002456 schistosoma
41	240.5	8.8	1929	13	098706	098706 notothenia
42	240.5	8.8	1941	13	09DGM4	09DGM4 gallus gall
43	240.5	8.8	2029	4	09C014	09C014 homo sapien
44	240.5	8.8	2069	4	09C013	09C013 homo sapien
45	240.5	8.8	2073	4	09C012	09C012 homo sapien

## ALIGNMENTS

RESULT 1	ID	PRELIMINARY	PRT	414 AA.
060527	060527			
AC	060527			
DT	01-AUG-1998 (TREMBlrel. 07, Created)			
DT	01-AUG-1998 (TREMBlrel. 07, Last sequence update)			
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)			
DE	Antigen NY-CO-8 (Fragment).			
GN	NY-CO-8.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=COLORECTAL CARCINOMA;			
RX	MEDLINE=98272252; PubMed=9610721;			
RA	Scanlan M.J., Chen Y.T., Williamson B., Gure A.O., Stockert E.,			
RA	Gordan J.D., Tureci O., Sahin U., Pfeundschn M., Old L.J.;			
RT	Characterization of human colon cancer antigens recognized by			
RT	autologous antibodies. "			
RL	Int. J. Cancer 76:652-658(1998).			
DR	EMBL; AF039690; AAC18039.1; -.			
FT	NON_TER			
SO	SEQUENCE	414 AA;	48616 MW;	69140CBF04406D7A CRC64;

## Alignment Scores:

Pred. No.:	8,58e-124	Length:	414
Score:	2062.00	Matches:	414
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	75.48%	Indels:	0
DB:	4	Gaps:	0

US-09-502-945-1 (1-1552) x 060527 (1-414)

```

OY 1 CTTTGAGATGATCCGAGAACTTAATCTTATGAGAAAAGTGTGAATTCAGAA 60
    |||||||
DB 1 LeuLeuAspAlaSerGlnLysLeuLysLeuThrTyrGlnGlnLysCysGlnLysGln 20
OY 61 TCCCAATTAAGCTTTTGAGAACGACTTACGTAATATCAGAGACTTGTGAAGTCTT 120
    |||||||
DB 21 SerGlnLeuLysPheLeuLysAsnAspLeuAlaGlnTyrGlnThrCysGlnAspLeu 40
OY 121 AAAGAGCACTAAAGCATTAAGAAATTTCTTGCGCTTAATACCTTGAACGCTGTGGT 180
    |||||||
DB 41 LysGlnGlnLeuLysHisLysGlnPheLeuLeuAlaAlaAsnThrCysAsnArgValGly 60
OY 181 GGTCTTTGTTTAAATGTGCTCAGCATGAGCTGTTCTTCCCAACCCATCTAATGTT 240
    |||||||
DB 61 GlyLeuCysLeuLysCysAlaGlnHisGlnAlaValLeuSerGlnThrHisThrAsnVal 80
OY 241 CATATGACAGACCATGCAAGAGCTGTTAAGAAGAGATGACTGTATGTCGACATGATT 300
    |||||||
DB 81 HisMetGlnThrIleGlnLysArgLeuValLysGlnLysAspAspLeuSerAlaLeuVal 100
OY 301 TCCGTAAGAGAGAGCTTGCGAGATACGACGAGCAAGAGCAAGCTTATGAACAGTGG 360
    |||||||
DB 101 SerValArgSerSerLeuLysAspThrGlnGlnArgGlnAlaSerAlaTyrGlnGlnVal 120
OY 361 AAACAAGTTTTCCAATATCTGAGAGAGCCATTTTGAAAAACCAAGCTTATTCGAG 420
    |||||||
DB 121 LysGlnValLeuGlnIleSerGlnGlnAlaAsnPheGlnLysThrLysAlaLeuIleGln 140
OY 421 TGTGACAGATTGAGAGAGAGCTGAGAGAGCGAGCGAGCTTGAAAGAACTTGA 480
    |||||||
DB 141 CysAspGlnLeuLysArgLysLeuGlnLysArgGlnAlaGlnLysArgLysLeuVal 160
OY 481 TCTCAGCAAGAGAAAGGCCATTGAGAAACATGATGATGAAAAAGAAATTAACGAA 540
    |||||||
DB 161 SerGlnGlnGlnLysArgAlaIleGlnLysAspMetLeuLysGlnLysThrLysGln 180
OY 541 AGGGAGTACATGGAGATCAAGAGATGTCGTCACAGAAATTCGCCAAGCTGAGCGCC 600
    |||||||
DB 181 ArgGlnLysArgLysSerLysMetLeuIleLeuSerGlnAsnIleAlaGlnLeuGlnAla 200
OY 601 CAGGTGAAAAGAGTTTAAAGAGAAAGATTTGACATTAATCAATGAGAGAAATTCAA 660
    |||||||
DB 201 GlnValGlnLysValThrLysGlnLysIleSerAlaIleAsnGlnLeuGlnIleGln 220
OY 661 AGCCAGCTGGCTTCGCGGAATTCGATGTCACAAAGCTGTGGAGAAATTCGCTATCAG 720
    |||||||
DB 221 SerGlnLeuAlaSerArgLysIleMetAspValThrLysValCysGlyGlnMetArgTyrGln 240
OY 721 CTGATTAACCAATGAGATGAGAGATGAGAGAAAGAGACACAGATGTCAGAGCA 780
    |||||||
DB 241 LeuAsnLysThrAsnMetCysLysAspGlnAlaGlnLysGlnHisArgLysIlePheArgAla 260
OY 781 AAAACTAACAGGAGATCTTGAATTAAGATCAGAAATGAGAAATGAGAAATGAGAACTG 840
    |||||||
DB 261 LysThrAsnArgAspLeuGlnLysAspGlnGlnIleGlnLysLeuArgLysIleGlnLeu 280
OY 841 GATGAAGCAAAACACTTGGAAACAGAGACAGACAGAGCCCTGGCCAGAGAGAGAG 900
    |||||||
DB 281 AspGlnSerLysGlnHisLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
OY 901 TGCCTGAGAGCTAAAGAAATGCTGGCGCAATCTGAGACCAACTGCACTCCACAGATCT 960
    |||||||

```

```

DB 301 CysLeuArgLeuThrGlnLeuLeuGlnLysSerGlnHisGlnLeuHisLeuThrArgSer 320
OY 961 GAAATAGCTCAACTCAGTCAAGAAAAAGTATATCATATGATTAATTTGGAAAAATTACAG 1020
    |||||||
DB 321 GlnIleAlaGlnLeuSerGlnGlnLysArgTyrThrTyrAspLysLeuGlnLysLeuGln 340
OY 1021 AGAAGAAATGAAAGAAATTTGGAGACAGAGTGTCTCCAGAGAGCAGAGCTTGCAGAGAGTGCAG 1080
    |||||||
DB 341 ArgArgAsnGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 360
OY 1081 AAGCAAGGCTTAAGCAGCTGTGATAGCAGACAGCCAGCCAGCCAGCCAGCTGTGTCAG 1140
    |||||||
DB 361 LysGlnArgLeuArgGlnLeuAspLysHisSerGlnAlaThrAlaGlnLeuValGln 380
OY 1141 CTCCTCAGACAGACAGAACCCAGCTTCTCTGAGAGAGCAGAGCCTTGCAGAGAGTGCAG 1200
    |||||||
DB 381 LeuLeuSerLysGlnAsnGlnLeuLeuGlnArgGlnSerLeuSerGlnLysValAsp 400
OY 1201 CGGCTGGGAGCCAGTTACCCAGCATGCAACATCTGATTCG 1242
    |||||||
DB 401 ArgLeuArgThrGlnLeuProSerMetProGlnSerAspCys 414

```

RESULT 2

Q9CSM9 PRELIMINARY; PRT; 468 AA.

AC Q9CSM9;

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE 2700048621Rik protein (Fragment).

GN 2700048621Rik.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=EMBRYO;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi T., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamaneke I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H., Kuehl P., Lewis S., Matsuo Y., Nikolaio I., Pesole G., Queckenbush J., Schriml L.M., Stanbly F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustlinclch S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyono-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S., Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:685-690(2001).

DR EMBL: AK012392; BAB28209.1; -

DR MGD: MGI:1924066; 2700048621Rik.

FT NON-TER 468

SO SEQUENCE 468 AA; 53462 MW; DABECB46808DFBD4 CRC64;

## Alignment Scores:

Pred. No.:	6.61e-52	Length:	468
Score:	935.00	Matches:	184
Percent Similarity:	89.54%	Conservative:	30
Best Local Similarity:	76.99%	Mismatches:	25
Query Match:	34.22%	Indels:	0
DB:	11	Gaps:	0

US-09-502-945-1 (1-1552) x Q9CSM9 (1-468)

[illegible]

Pred. No.:	6.58e-46	Length:	393
Score:	841.00	Matches:	166
Percent Similarity:	100.00%	Conservative:	0
Best local Similarity:	100.00%	Mismatches:	0
Query Match:	30.78%	Indels:	0
DB:	4	Gaps:	0

US-09-502-945-1 (1-1552) x Q9P0F1 (1-393)

QY	16	GAGAGCTAAACTTACTTCCTTATGAGGAAAAGTGGAAATGTGAGGAAATCCCAATTGAAGTTT	75
Db	226	GIUUSLEUUSLEUHTTIRYRIGUIGUUSYCSYSGUIIEGUGIUSERGINSLEUUSPHE	24
QY	76	TTGAGCAACGACTTACCTGAATATCAGAGAACTGTGAGATCTTAAAGACAACTAAAG	131
Db	246	LEUARGSNAPSPLEUHLAIGIURYGIRNGIRTHRCYSGIUSAPLEUUSYSGIUGINSLEUUS	26
QY	136	CATPAAGAAATTTCTTGGCTGCTTAATACCTGTGAACCGTGTGGTGCTCTTTGTTGAAA	19
Db	266	HISYSGIURPHEUENLEUHLAIAASNTHRCYSANARGVALIGIYGLIEUCYSLEUUS	28
QY	196	TGTCTCAGCATGAACCTGTCTTCTTCCCAAAACCCATCTATGTTCATGTATGACAGACATC	25
Db	286	CYSALAEIHLISIGUHLAVALLEUSERGINTHTHSTRASNVAHLHSMETGINTHTLIE	30
QY	256	GAAGAAGATGGTTTAAAGAAAGACATGACTGTGATGTCGTGCACAGATTTCCGTAAAGACACG	31
Db	306	GIUNARGLEUVALIYSGIUNARGSPSPLEUHNETSERALALEUVALISERVALARGSERSER	32
QY	316	TTGGCAGATACGACGAGAAAAGACGAAGTGGCTTATGAAACAGTGAACAGTGAACAATTTTGCAA	37
Db	326	LEUHLASPHRCLINGINSIRNGIUNLASERLATRYGUGINGUVALIYSGIUNLAEUGIN	34
QY	376	ATATCTGAGAGAACCCAAATTTGTAAGAAAACCAAGCGTTTAAATCCAGTGTGACGCTTGAGG	43
Db	346	ILSERIIGUHLIUNLAASNHEGILUUSYTHRYSLALEUHLIEGINSYASPGINSLEUARG	36
QY	436	AAGAGCTGGAGAGGACGACGACGAGCATTTGTAAGAAAAGAACTTGCAATTCAGCAAGAGAAA	49
Db	366	LYSLIUENSIGUNARGIUNHLAIGIUNARGLEUGIUSYSGIULEUHLASERGINGLINGIULYS	38
QY	496	AGGGCCATTGAGAAAGACATGATG 519	
Db	386	ARGALALIEGILUUSASPMETMET 393	
RESULT 4			
QY	Q8R3KO	PRELIMINARY;	PRT; 379 AA.
AC	Q8R3KO;		
DT	01-JUN-2002 (TREMBLrel. 21, Created)		
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)		
DE	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
Df	Similar to RIKEN CDNA 2700048G21 gene.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
CX	NCBI_TaxId=10090;		
RN	[1]		
RA	SEQUENCE FROM N.A.		
RA	Strasberg R.;		
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: BC025138; AAH25138.1; ..		
QO	SEQUENCE 379 AA; 42775 MW; B20D32A9E193F53C CRC64;		

### Alignment Scores:

Pred. No.:	1,7e-25	Length:	379
Score:	521.00	Matches:	102
Percent Similarity:	88.55%	Conservative:	14
Best local Similarity:	77.86%	Mismatches:	15
Query Match:	19.07%	Indels:	0
DB:	11	Gaps:	0

### Alignment Scores:

US-09-502-945-1 (1-1552) x Q8R3K0 (1-379)

```

OY 16 GAGAACTAAACTTACTTATGAGAAAGTGTGAATTTGAGGATCCCAATGAAGTTT 75
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 192 GUAAGLeuLysLeuThrTyrglnAlaLysThrAspLeuLysGlnLeuMetLeu 211
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 76 TTGAGAACGACTTACCTGAATATCAGAGAACTTGTGAAGATCTTAAAGACAACTAAAG 135
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 212 LeuArgLysLeuLeuAlaLysThrCysGlnAspLeuLysGlnLeuMetLeu 231
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 136 CATTAAGAAATTTCTGTGCTGCTAATCTTGAACCGCTTGGCTGCTTGTGAAA 195
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 232 HisLysGlnSerLeuLeuAlaSerAlaSerSerArgValGlyLysCysLeuLys 251
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 196 TGTGCTCAGCATGAAAGCTGTCTTCCCAACCATACATAATGTTATGACAGACATC 255
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 252 CysAlaGlnHisGlnAlaValLeuSerGlnThrHisSerAsnValHisIleGlnThrIle 271
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 256 GAAAGACTGTTAAAGAAAGACATGCTGATGCTGACCTAGTTCCCTAAGACAGC 315
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 272 GUAAGLeuThrLysGlnArgAspLeuMetSerValLeuValSerValArgSerSer 291
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 316 TTGGCAGATACGACGACCAAGACAGACAGCTTATGACAGCTGAACAGTTTGCAG 375
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 292 LeuAlaGlnHisGlnAlaLysThrArgGlnThrSerAlaTyrglnGlnValHisAlaValGln 311
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 376 ATATCTGAGGAAGCCCAATTTTGAACCAAG 408
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 312 MetThrGlnGlnAlaLysAspMetGlnThrLys 322
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

## RESULT 5

O9CYHO PRELIMINARY: PRT: 314 AA.

AC O9CYHO;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)  
 DE 5730470G24RIK.  
 GN 5730470G24RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;

RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,  
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
 RA Schiml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyokawa K., Wang K.H., Weltz C., Whitaker C., Wilming L.,  
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection";  
 RL Nature 409:685-690(2001).  
 DR EMBL; AK017691; BAB30877.1;  
 DR MGD; MGI:1917837; 5730470G24RIK.  
 SQ SEQUENCE 314 AA: 35730 MW; B41F3B08B8E418F CRC64;

## Alignment Scores:

Pred. No.: 3-87e-14 Length: 314  
 Score: 343.00 Matches: 66

Percent Similarity: 88.24% Conservative: 9  
 Best Local Similarity: 77.65% Mismatches: 10  
 Query Match: 12.55% Indels: 0  
 DB: 11 Gaps: 0

US-09-502-945-1 (1-1552) x O9CYHO (1-314)

```

OY 16 GAGAACTAAACTTACTTATGAGAAAGTGTGAATTTGAGGATCCCAATGAAGTTT 75
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 230 GUAAGLeuLysLeuThrTyrglnAlaLysThrAspLeuLysGlnLeuMetLeu 249
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 76 TTGAGAACGACTTACCTGAATATCAGAGAACTTGTGAAGATCTTAAAGACAACTAAAG 135
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 250 LeuArgLysAspLeuAlaLysThrCysGlnAspLeuLysGlnLeuMetLeu 269
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 136 CATTAAGAAATTTCTGTGCTGCTAATCTTGAACCGCTTGGCTGCTTGTGAAA 195
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 270 HisLysGlnSerLeuLeuAlaSerAlaSerSerArgValGlyLysCysLeuLys 289
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 196 TGTGCTCAGCATGAAAGCTGTCTTCCCAACCATACATAATGTTATGACAGACATC 255
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 290 CysAlaGlnHisGlnAlaValLeuSerGlnThrHisSerAsnValHisIleGlnThrIle 309
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
OY 256 GAAAGACTGTTAAAG 270
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 310 GUAAGLeuThrLys 314
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

## RESULT 6

O76329 PRELIMINARY: PRT: 1738 AA.

AC O76329;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)  
 DE Interactin.  
 GN ABPD.  
 OS Dictyostellium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyosteliales; Dictyostellium.  
 OX NCBI\_TaxID=44689;

RP SEQUENCE FROM N.A.  
 RX MEDLINE=98365468; PubMed=9700162;  
 RA Rivero F.J., Kuspa A., Brokamp R., Matzner M., Noegel A.A.;  
 RT "Interactin, an actin-binding protein of the alpha-actinin superfamily  
 in Dictyostellium discoideum, is developmentally and cAMP-regulated and  
 associates with intracellular membrane compartments.";  
 RL J. Cell Biol. 142:735-750(1998).  
 DR EMBL; AF057019; AAC34582.1;  
 DR HSP; P46939; IQAG.  
 DR InterPro: IPR001589; Actbind\_actin.  
 DR InterPro: IPR001715; Calponin-like.  
 DR InterPro: IPR001990; Grabin.  
 DR InterPro: IPR001451; Hexapep\_transf.  
 DR InterPro: IPR002017; Spectrin.  
 DR Pfam: PR00307; CH\_2.  
 DR SMART; SM00033; CH\_2.  
 DR PROSITE; PS00019; ACTININ\_1; UNKNOWN\_1.  
 DR PROSITE; PS00020; ACTININ\_2; 1.  
 DR PROSITE; PS00021; CH\_2.  
 DR PROSITE; PS00422; GRANINS\_1; UNKNOWN\_1.  
 DR PROSITE; PS00101; HEXAPEP\_TRANSFERASES; UNKNOWN\_1.  
 SQ SEQUENCE 1738 AA: 204427 MW; 577A99D2EC79AF5C CRC64;

## Alignment Scores:

Pred. No.: 1.15e-09 Length: 1738  
 Score: 272.50 Matches: 106  
 Percent Similarity: 42.16% Conservative: 109  
 Best Local Similarity: 20.78% Mismatches: 162  
 Query Match: 9.97% Indels: 133  
 DB: 5 Gaps: 17

US-09-502-945-1 (1-1552) x O76329 (1-1738)





Db	1003	Ilkasper	.....	....		MetSerGlnGluSerGlu	1011
Oy	235	AATGTCATATGCAG	-----	-----	ACCATGAAAGACTGGTTAAAGAAAGA	276	
Db	1012	AsnPhcGlnIleGluArgGlySerIleGluYpsAsnIleGluGlnLeuYpsThrIle	.....		.....	1031	
Oy	277	GATACATGATGTCGCACATGATTCCGTAGAGCAGCGTGGCAGATACGACGACAAAGA	336				
Db	1032	SerAspLeuGlnGlnThrLysGluGluIleIleSerLysSerAspSerSerLysAspGlu	1051				
Oy	337	GAACACAGTGGCTTATGACACAGGTGAACAAAGTTTGTCCAAATATCTCGAGAGACCAATTT	396				
Db	1052	TyrGlnSerGlnIleSerLeuLeuYpsGluYpsGluGlnGluThrAlaThrThrAlaAsnAsp	1071				
Oy	397	GAAGAAACCAAGCCTTTAATCCAGCTGACACCTTAGACAGAGCTGGACAGCGAC	453				
Db	1072	GluAsnValAsnYpsIleSerGluLeuThrLysThrArgGluGluLeuGluGluLeu	1091				
Oy	454	-----	-----	GGGAGGCGACTT	465		
Db	1092	AlaAlaTyrLysAsnLeuLysAsnGluLeuGluThrLysLeuGluThrSerIleYpsAla	1111				
Oy	466	GAAGAAAGCACTGGCATTCACGACAAAGAGAAAGGGCCATTGAGAAAGACATGATGAAAG	525				
Db	1112	LeuYpsGluValLysGluAsnGlnGluGlnHisLeuLysGluGluYpsIleGlnLeuGluYps	1131				
Oy	526	GAATTAACGAAGAAAGGAGTACAGTGGATCCAAAGATGTGATCTGTCTCAGAAATTT	585				
Db	1132	GluAlaThrGluThrLysGlnGlnLeuAsnSer	1142				
Oy	586	GCCACACGGAGCGCCGAGGGGAGAAAGGTTACAAAGAAAGATTTCAGCTTATTAACAA	645				
Db	1143	-----LeuArgAlaAsnLeuGluSerLeuGluYpsGluHisGluAspLeuAlaAlaGln	1160				
Oy	646	CTGAGAGCAATTCACAAACCAGCTGGCTTCGCGGAATGATGATCCAAAGCGTGTGGA	705				
Db	1161	LeuYpsLysTyrGlnGlnGlnIleAlaAsnLysGluAlaGlnGlnTyrAsnGluGluIleSer	1180				
Oy	706	GAATTCGGCTATCAGCTGAATTAACCAACATGGAG	744				
Db	1181	GlnLeuAsnAspGluIleThrSerThrGlnGlnGluAsnGluSerIleLysLysLysAsn	1200				
Oy	745	GATGAGGCGAAGAAAGGACGACAGAGGTTCCAGCAAAACATACAGGATCTTGAATTT	804				
Db	1201	AspGluLeuGlnGluYpsValLysAlaMetLysSerThrSerGlnGlnGlnSerAsnLeu	1220				
Oy	805	AAAGATCAGCAATATAGAAATTTAGAAATTTAGAACTGGATGGAAGCAAAACACTTTGGAA	864				
Db	1221	LysLysSerGlnIleAspAlaLeuAsnLeuGlnIleLysGluLeuYpsLys	1238				
Oy	865	CAGGAGCAGCAGAAAGCAGCCCTG	906				
Db	1239	AsnGlnThrAsnGlnAlaSerLeuLeuGlnGlnIleLysSerValGlnSerGlnThrVal	1258				
Oy	907	AGACTAACACAAACCTCGTGGCGCAATCTGAGACCAACTGCACCTACACAGATCTGAATA	966				
Db	1259	LysIleLysGluLeu	1274				
Oy	967	GCTCACTCAGTCAAGAAAGGTTATACATATGATTAATTTGGAAAG	1017				
Db	1275	SerIleuGlnAspLysLysLeuYpsAlaSerGluAspLysAsnSerLysTyrLeuGlnLeu	1294				
Oy	1018	CAGGAGAAATGAGAAATTTGGAGAAACAGTGTCTCCACATAGGAGAGTACATGACAGC	1077				
Db	1295	GlnYpsGluSerGluYpsIleLysGlnGlnLeuAspAlaLysThrThrGluLeuYpsIle	1314				
Oy	1078	ATGAGCAAAAGCTTAAGGACGCTGATTAAGCACAGCCAGCCACACCCACAGCTGGTG	1137				
Db	1315	GlnLeuGlnLysIleThrAsnLeuSerLysAlaLysGlnLysSerGlnSerGlnLeuSer	1334				
Oy	1138	CAGTCTCTCAGCAAGCAAGCAAGCTTCTCTGAGAGGCAAGCGCTGTGGAAAGAGTG	1197				

Db	1335	ArgleuylslysthrSerSerglu-----	GlualglysaAlaIuGlIuGlInleu	1351
Qy	1198	GACCGGCTGGCGGACCGACGTTA	1218	
Db	1352	GlulysleuylsaAngluIle	1358	
RESULT 8				
	Q94992	PRELIMINARY;	PRT; 2017 AA.	
AC	Q94992; Q24138;			
DT	01-FEB-1997 (TREMBLrel_02, Created)			
DT	01-FEB-1997 (TREMBLrel_02, last sequence update)			
DT	01-JUN-2002 (TREMBLrel_21, last annotation update)			
DE	Nonmuscle myosin-II heavy chain.			
GN	ZIP OR CG15792.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_Taxid=7227;			
RX	[1]			
RE	SEQUENCE FROM N.A.			
RX	MEDLINE=6144835; PubMed=8568878;			
RA	Mansfield S.G., al-Shirawi D.Y., Ketchum A.S., Newbern E.C.,			
RA	Richard D.P.;			
RT	"Molecular organization and alternative splicing in zipper, the gene			
RT	that encodes the Drosophila non-muscle myosin II heavy chain.";			
RL	J. Mol. Biol. 255:98-109(1996).			
DR	EMBL; U35816; AAB09048.1; -.			
DR	EMBL; U35816; AAB09051.1; -.			
DR	HSSP; P10587; IBR2.			
DR	FlyBase; FBgn0005634; zip.			
DR	InterPro; IPR0000048; IQ_region.			
DR	InterPro; IPR001609; myosin_head.			
DR	InterPro; IPR004009; myosin_N.			
DR	InterPro; IPR02928; myosin_tail.			
DR	Pfam; PF00612; IQ; 1.			
DR	Pfam; PF00063; myosin_head; 1.			
DR	Pfam; PF02736; Myosin_N; 1.			
DR	Pfam; PF01576; Myosin_tail; 1.			
DR	PRINTS; PR00193; MYOSINHEAVY.			
DR	ProDom; PD000355; myosin_head; 1.			
DR	SMART; SM00015; IQ; 1.			
DR	SMART; SM00242; MYSC; 1.			
DR	PROSITE; PS50096; IQ; 1.			
DR	SEQUENCE 2017 AA; 232089 MW; 8C76FEFZEBD02EBE CRC64;			
Alignment Scores:				
	Pred. No.:	1.73e-08	Length:	2017
	Score:	254.00	Matches:	110
	Percent Similarity:	43.15%	Conservative:	101
	Best local Similarity:	22.49%	Mismatches:	174
	Query Match:	9.30%	Indels:	104
	Db:	5	Gaps:	18
US-09-502-945-1 (1-1552) x Q94992 (1-2017)				
Qy	4	CTGATGTCATCCCGAAGCACTTAATGAGAAAGTGTGAATTGACATCC	63	
Db	1143	IleAspGIuSerAlaThrLysAlaThrAlaGluLysAlaGluIuSer	1162	
Qy	64	CAATTGAAGTTTGGAGAGCACTA-----	GCTGAATATCAG	102
Db	1163	GluIuAlaGluIleGluIuAspLeuGluAlaGluLysAlaAlaArgAlaLysAlaGlu	1182	
Qy	103	AGAACTGTGAAGATCTTAAGACCACTRAA-----	CATAAGAAATTTCTTGCGCT	156
Db	1183	LysValArgArgAspLeuSerGIuLeuGluAlaLeuLysAsnGIuIuLeuAspSer	1202	
Qy	157	GCTAATCTTGTAACCTGTGGTGGCTTTGTTGAATGTCGACAG-----	204	
Db	1203	LeuAspThrAlaAlaGluIuLeuAlaGlySerLysArgGIuGluIuAlaThr	1222	

Qy 205 -----CATGAAGCTGTTCTT----- 219  
Db 1223 LeuLysLysSerLeuGluGluThrValAsnHisGluGluValLeuAlaAspMetArg 1242  
Qy 220 -----TCCCAAAACCCTACTAATGTCATATGACACCATTGAAAGACTGGTTAA 270  
Db 1243 HisLysHisSerGlnLeuAsnSerIleAsnAspGln-----LeuGluAsnLeuArgLys 1261  
Qy 271 -----GAAAGAGATGACTGTATG 288  
Db 1262 AlaLysThrValLeuGluLysAlaLysGlyThrLeuGluAlaGluAsnAlaAspLeuAla 1281  
Qy 289 TCTGCAGTACTTTCGTAAGAGACAGCTTGCGCATACGACAAAGAGAGACAGTCT 348  
Db 1282 ThrGluLeuArgSerValAsnSerSerArgGlnGluAsnAspArgArgArgLysGlnAla 1301  
Qy 349 TATGACAGGTGAAACAAGTTTGCATATCTGAGAGCCAAATTTGAAAAACCAAG 408  
Db 1302 GluSerGlnIleAlaGlu-----LeuGlnVal-----LysLeuAlaGluIleGluArgAlaArg 1319  
Qy 409 GCTTTAATCCAG-----TGTGACAGTTGAGGAAGGCGGAGGAGCGGAGCGCA 462  
Db 1320 SerGluLeuGlnGluLysCysThrLysLeuGlnGlnGluAlaGluAsnIleThrAsnGln 1339  
Qy 463 CTTGAAAA-----GAAGTTCATCTCAGCAAGAGAAA 495  
Db 1340 LeuGluGluAlaGluLeuLysAlaSerAlaValLysSerAlaSerAsnMetGluSer 1359  
Qy 496 AGGCGCATTTGGAAGACATGATGAAAAAGAAATACGAAAGAAAGGAGTACATGGGA 555  
Db 1360 GlnLeuThrGluAlaGlnGlnLeuGlnGluGlnGluThrArgGlnLys----- 1375  
Qy 556 TCAGAAAGTGTGATCTGTCAGAAATATGCCCACTGGAGGCGCCAGCTGAAAAAGT 615  
Db 1376 -----LeuGlyLeuSerSerLysLeuArgGlnIleGluSerGlnLysGlnAlaLeu 1392  
Qy 616 ACAAGAGAAAAAGATTTCAGCTATTATCACTGAGAAATTCAAAGCCAGCTGGCTCT 675  
Db 1393 GlnGlu-----GlnLeuGlnGluAspArgGluAlaLysArgAsn 1405  
Qy 676 CGGGAATATGATGTCACAAAGGTGTGTGAGAAATGCCCTTACAGCTGAATTAACCAAC 735  
Db 1406 TyrGluArgLysLeuAlaGlnValThrThrGlnMetGln-----GluIleLysLysLysAla 1424  
Qy 736 ATGGAAGATGAGGCGAGAAAGGACAGACAGAGTTCCAGACAAAAATCAACAGGAT 795  
Db 1425 GluGluAspAlaAspLeuAlaLysGlnLeuGlnGluGlnLysLysArgLysLeuAsnLysAsp 1444  
Qy 796 CTTGAAATTAAGATCAGAAATAGAGAAATTCAGAAATAGAA-----CTGATGAA 846  
Db 1445 IlleGluAlaLeuGlnArgGlnValLysLysGlnLeuIleAlaGlnAsnAspArgLysAspLys 1464  
Qy 847 AGCAACACACCTTGGAACAGAGACAGACAGAGCC----- 885  
Db 1465 SerLysLysLysIleGlnSerGlnLeuLysAlaThrIleGluLeuGluAlaGlnArg 1484  
Qy 886 -----CTGGCCAGAGAG 897  
Db 1485 ThrLysValLeuGluLeuGlnLysLysGlnLysAsnPheAspLysIleLeuAlaGlnGlu 1504  
Qy 898 GAGTGCCCTG-----AGACTAAGAACTGCTGGCGCAATTCAGACCAACTGCACCTC 951  
Db 1505 LysAlaIleSerGlnGlnIleAlaGlnGlnArgAspThrAlaGlnArgGlnAlaArgGln 1524  
Qy 952 ACCAGATCTGAATAGCTCACTCAGTCACAGAAAAAGGTTACATATGATTAATGGGA 1011  
Db 1525 LysGluThrLysValLeuSerValSerArgLysGluLeuAspGluAlaPheAspLysIleGlu 1544  
Qy 1012 AAGTTACAGACAGAAATGAGATTCGAGACAGCTGTGCCAGATGGAGAGTACAT 1071  
Db 1545 AspLeuGluAsnLysArgLysThrLeuGlnAsnGlnLeuAspAspLeuAlaAsnThrGln 1564  
Qy 1072 GAGACGATGAAGCAAGGCTTAAGCAGCTGATTAAGCACAGCCAGGCCACAGCCAGCAG 1131

Db 1565 GlyThrAlaAspLysAsnValHisGlnLeuGlnLysAlaLysArgAlaLeuGluSerGln 1584  
Qy 1132 CTGGTGCAGCTCTTCAGACAGACAGACAGCTTCTCGTGGAGAGGACGCTGTGGAA 1191  
Db 1585 LeuAlaGlu-----LeuLysAlaGlnAsnGlnGluLeuGlnAspAspLeuGlnLeuThrGln 1603  
Qy 1192 GAGTGCAGCGGCTGGAGCCAGCTTA 1218  
Db 1604 AspAla-----LysLeuArgLeuGluVal 1611  
RESULT 9  
Q9W0W8 PRELIMINARY; PRT; 2056 AA.  
AC Q9W0W8: 01-MAY-2000 (TReMBLrel. 13, Created)  
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE ZIP protein.  
GN ZIP OR CG15792.  
OS Drosophila melanogaster (fruit fly).  
OC Euarthropoda; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferreira C., Fertiza S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegyam C.,  
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod K.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Syvaskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveli J.S., Zhan M., Zhan G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster";  
RL EMBL: AE003465; AAF47311.1; .  
DR HSSP: P10587; IBR2.  
DR FlyBase; FBgn0005634; zip.  
DR InterPro; IPR000046; IQ\_region.  
DR InterPro; IPR001609; myosin\_head.  
DR InterPro; IPR004009; myosin\_N.





RT gene expressed during embryonic skeletal muscle fiber formation.  
 RL J. Biol. Chem. 271:17047-17056(1996).  
 DR EMBL: U53862; AAC59912.1; -  
 DR EMBL: U53861; AAC59911.1; -  
 DR HSSP: P08799; 1MND.  
 DR InterPro: IPR000048; IQ\_region.  
 DR InterPro: IPR001609; myosin\_head.  
 DR InterPro: IPR004009; Myosin\_N.  
 DR InterPro: IPR002928; Myosin\_tail.  
 DR InterPro: IPR002017; Spectrin.  
 DR Pfam: PF00612; IQ\_2.  
 DR Pfam: PF00063; myosin\_head\_1.  
 DR Pfam: PF02736; Myosin\_N\_1.  
 DR Pfam: PF01576; Myosin\_tail\_1.  
 DR PRINTS: PR00193; MYOSINHEAVY.  
 DR ProDom: PD000355; myosin\_head\_1.  
 DR SMART: SM00015; IQ\_1.  
 DR SMART: SM00242; MTC\_1.  
 DR PROSITE: PS50096; IQ\_1.  
 DR SEQUENCE 1931 AA; 221745 MW; D2C1SDDE9BC8DE9 CRC64;

## Alignment Scores:

Pred. No.:	1 87e-08	Length:	1931
Score:	253.50	Matches:	123
Percent Similarity:	39.66%	Conservative:	86
Best Local Similarity:	23.34%	Mismatches:	181
Query Match:	9.28%	Indels:	137
DB:	13	Gaps:	20

US-09-502-945-1 (1-1552) x Q91973 (1-1931)

QY 19 AACCTAACTTACTTATGAGAAAGTGTGAATTCAGCAATCCCAATGAACTTTTTC 78  
 DB 1235 LYLALALYALALASINLEUGLULYSMETCYARGSERTHRGILNUSPGLMETASGLNHS 1254  
 QY 79 AGAAGAGACTTACTGAAATTCAGCAACTCTTGGAAGAT 117  
 DB 1255 ARGASNLYSLEULUGLSERGLNARGTHVALTHRASPLEUSERTHRGINARGALALYS 1274  
 QY 118 -----CTTAAAGACACTTAAGCATTAAGCATTAAGCATTTCTTCG 153  
 DB 1275 LEUGLTHRGILNUSNERSERGLULSERARGGLNLEUGLULYSGLIALAPHEILEASN 1294  
 QY 153 ----- 153  
 DB 1295 GINLEUMETARGSLYLSLEUTHRYTHRGINGLNLNLEUGLUSPLEULYSARGGLNLEU 1314  
 QY 154 -----GCTGCTAATCTTACTTAACCGCTGTGGCTGCTTTGTAATGCTGCACGAT 207  
 DB 1315 GLUGLUGLUALALYALALASASNALALEUALAHIS--ALALEUGINSERIALAGLNHS 1333  
 QY 208 GAAGCTGTTCTTCCCAACCCACTACTAATGTCATATGACAGACCACTGAAGACTGGTT 267  
 DB 1334 ASPCYASPLEULNARG-----GLUGLNYRGILUGLUGLUMET 1346  
 QY 268 AAGAAGAGAGACTGATGCTGCACACTGTTCCGTAAAGACAGCTTGCA----- 321  
 DB 1347 GLUALALYALALAGLULEUGLNARGALALEUSERLYSALAAASNERSERGLUALALAGINTRP 1366  
 QY 322 -----GATACGACGAAAGAGAA-----GCAAGTGGCTTTGAA 354  
 DB 1367 ARGTHLYSTYRGILUTHRASPALALIEGLNARGTHRGILUGLULEUGLUGLUALALYLS 1386  
 QY 355 CAGGTGAACAGTTTTCGCAATATCTGAGAGAGCC----- 390  
 DB 1387 LYSLEUALAGLNRGLEUGLNGLUALAGLUGLUALAVALGUALAVALASNALALYSCYS 1406  
 QY 391 ---AATTTGAAAAAACCAAGCTTTAATCCAG----- 420  
 DB 1407 SERSERLEUGLULYSTRHYSHISARGLEUGLNASGLULIEGLUSPLEUMETALASP 1426  
 QY 421 -----TGTGACACAG 429  
 |||:::

DB 1427 VALGILNARGSERASNALAALALALALALEUASPLYSGLINARGASNPLEASPLYS 1446  
 QY 430 TTGAGCAAGAGAGCTGCGAGAGAGCGAGCGACTTGTAAGAAAGAACTT---GCATCTCAG 486  
 DB 1447 ILEULEUSERLUTRPRYSGLINLSPHEGLUGLSERGLNTHRGILNLEUGLUALASERGLN 1466  
 QY 487 CAAGAGAAAGGGCCATTTGAGAAAGACATGATGAAA-----AAGGAATA 531  
 DB 1467 LYSGLIUALARGSERLEUSERTHRGILNLEUPHEULYSASNALATYRGILUGLUSER 1486  
 QY 532 ACGAAGAAAGAGAGATCATGCGATCAAGATGTGATCTTGCTCGCAATATGCCCA 591  
 DB 1487 LEUGLNIHISLEULNTHRPHEULYSARGGLUASNLYSASNLEUGLNGLUGLILSERASP 1506  
 QY 592 CTGAGAGCCCAAGGTGAGAAAGTTACAAAGAAAGATTTCAGTATTAATCACTGAG 651  
 DB 1507 LEUTHRGILNLEUGLUALASERGLNLS-----SERILEHISGLULEUGL 1522  
 QY 652 GAAATTCMAAGCCAGCTGCTCGGAAATGATGTCMAAAGGTGTGGAAGAAATG 711  
 DB 1523 LYSVALARGYSGILNLEUASPALAGLULYSLEUGLULEUGLINALALALEUGLUALA 1542  
 QY 712 CGCTATCAGCTGATATAAACCAACATGCAAGAGATGAGCAGCAAAAAGCAGCAGAG 771  
 DB 1543 GLUALASERLEUGLNIHISGLUGLULYSILEUARGALAGLNLNLEUGLUPHEASGLN 1562  
 QY 772 TTCAGAGCAAAACTAAGACAGGATCTTGAATTAAGATGAGAAATAGAGAA----- 825  
 DB 1563 VALLYSALAGLUTYRGILUARGYLSLEUALAGLULYSASPLUGLULMETGLUGLINSER 1582  
 QY 826 -----TTGAGAAATAGAACTGATGAAAGCAACACACTTGGACAGCAGCAG 876  
 DB 1583 ARGASNLISLEULARGVAL---VALASPSERLEUGLNIHRSERLEUASPALAGLUTHARG 1601  
 QY 877 AAGGACCCCTGCGCAGAGAGAGCTGCTGAGCACTAACAACATG-----CTG 924  
 DB 1602 -----SERARGASNLUALALEUARGLEULYSLSYSMETGLULYASPLEU 1617  
 QY 925 GCGCAATCTGAGCAGCCAACTG---CACCTCCAGATCTGAAATAGCTCAACTCAGTCAA 981  
 DB 1618 ASNLUMETGLULIEGLNLEUSERHISALASNARGTHRLALAGLUALAGLULYSGLN 1637  
 QY 982 GAAAA-----AGGTATACATATGATTAATTTGGAAAATTACAGACA 1023  
 DB 1638 VALLYSALALEUGLNGLYTRYLEULYSASPTHRGILNLEUGLNUASPSALVALARG 1657  
 QY 1024 AGAATGAGAAATTTGGAGAACAGTGT----- 1050  
 DB 1658 ALASNLGLUSPLEULYSGLUASNIIIEALILEVALGILNARGASNLNLEUGL 1677  
 QY 1051 -----GTCCAGCATGGAGAGATGACATGAGACGATGCAAGCAAGCAAGCTTAAGCAGCTGGAT 1104  
 DB 1678 SERGLULEUGLUGLULEUARGLAMEVALIGLNLNLSERGLNARGALARGLYLSLEUALA 1697  
 QY 1105 AAGCAGACCCAGGCCACAGCCACAGCTGCTGCTCAGCAAGCAAGAAC---CAG 1161  
 DB 1698 GLUGLNGLULEUTHRGILUALASERGLUARGVALGILNLEUENHISSERGILNANTRHSER 1717  
 QY 1162 CTTCTCCTGAGAGGAGAGAGCTGTCGGAAGGTGAGACCGGAGACCGAGCTTAACC 1221  
 DB 1718 LEULIEHISNLINLSYLSYSMETGLUALASPILESERGLNLEUGLNIHTRGILUALG 1737  
 QY 1222 AGCATGCCACAAATCTGATTC 1242  
 DB 1738 GLUALALIEGLN---GLUCYS 1743  
 RESULT 12  
 Q14789 PRELIMINARY; PRT; 3259 AA.  
 AC Q14789; Q14398;  
 DT 01-NOV-1996 (TREMREL. 01, Created)  
 DT 01-NOV-1996 (TREMREL. 01, Last sequence update)  
 DT 01-MAY-2000 (TREMREL. 13, Last annotation update)

DE GIANTIN (GCB372) (MACROGOLGIN) (Golgi autoantigen, golgin subfamily B,  
 1).  
 GN GOLGI.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9418728; PubMed=7511208;  
 RA Seelig H.P., Schrantz P., Schroeter H., Wiemann C., Griffiths G.,  
 RA Renz M.;  
 RT "Molecular genetic analyses of a 376-kilodalton Golgi complex membrane  
 RL Mol. Cell. Biol. 14:2564-2576(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94257116; PubMed=8198703;  
 RA Seelig H.P., Schrantz P., Schroeter H., Wiemann C., Griffiths G.,  
 RA Renz M.;  
 RT "Macrogolgin--a new 376 kD Golgi complex outer membrane protein as  
 RT target of antibodies in patients with rheumatic diseases and HIV  
 RT infections.";  
 RL J. Autoimmun. 7:67-91(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95100974; PubMed=7802676;  
 RA Sonba M., Misumi Y., Fujiwara T., Nishioke M., Iehara Y.;  
 RT "Molecular cloning and sequence analysis of a human 372-kDa protein  
 RT localized in the Golgi complex.";  
 RL Biochem. Biophys. Res. Commun. 205:1399-1408(1994).  
 CC -1- FUNCTION: MAY PARTICIPATE IN FORMING INTERCISTERNAL CROSS-BRIDGES  
 CC OF THE GOLGI COMPLEX.  
 CC -1- SUBUNIT: DISULFIDE-LINKED HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. GOLGI MEMBRANE.  
 CC -1- ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS MAY ARISE BY ALTERNATIVE  
 CC SPLICING.  
 CC -1- DISEASE: ANTIGEN IN CHRONIC RHEUMATOID ARTHRITIS AND IN THE  
 CC AUTOIMMUNE DISEASE SJOEGREN'S SYNDROME.  
 DR EMBL: X75304; CAAS3052.1; -;  
 DR EMBL: D25542; BAA05025.1; -;  
 DR MIM: 602500; -;  
 KW Golgi stack; Antigen; Coiled coil; Transmembrane;  
 KM Alternative splicing.  
 FT DOMAIN 1 3235 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 3236 3256 POTENTIAL.  
 FT DOMAIN 3257 3259 LUMENAL (POTENTIAL).  
 FT DOMAIN 48 110 COILED COIL (POTENTIAL).  
 FT DOMAIN 127 223 COILED COIL (POTENTIAL).  
 FT DOMAIN 238 448 COILED COIL (POTENTIAL).  
 FT DOMAIN 460 526 COILED COIL (POTENTIAL).  
 FT DOMAIN 545 593 COILED COIL (POTENTIAL).  
 FT DOMAIN 677 956 COILED COIL (POTENTIAL).  
 FT DOMAIN 969 1028 COILED COIL (POTENTIAL).  
 FT DOMAIN 1062 1128 COILED COIL (POTENTIAL).  
 FT DOMAIN 1154 1245 COILED COIL (POTENTIAL).  
 FT DOMAIN 1301 1779 COILED COIL (POTENTIAL).  
 FT DOMAIN 1828 2781 COILED COIL (POTENTIAL).  
 FT DOMAIN 2797 2857 COILED COIL (POTENTIAL).  
 FT DOMAIN 2872 2993 COILED COIL (POTENTIAL).  
 FT DOMAIN 3026 3102 COILED COIL (POTENTIAL).  
 FT DOMAIN 3133 3185 COILED COIL (POTENTIAL).  
 FT DOMAIN 2420 2423 POLY-GLU.  
 FT DOMAIN 2993 2996 POLY-SER.  
 FT VARSPLIC 1 39 MISSING (IN REF. 3).  
 FT VARSPLIC 215 215 A -> AQLSSM (IN REF. 3).  
 FT CONFLICT 1765 1765 D -> G (IN REF. 3).  
 FT CONFLICT 2950 2950 H -> D (IN REF. 3).  
 SQ SEQUENCE 3259 AA: 376075 MW: 60376A20DBA178BD CRC64:

Alignment Scores: 1.84e-08 Length: 3259  
 Pred. No.: 253.50 Matches: 119  
 Score:

Percent Similarity: 40.33% Conservative: 104  
 Best Local Similarity: 21.52% Mismatches: 179  
 Query Match: 9.28% Indels: 151  
 DB: 4 Gaps: 24  
 US-09-502-945-1 (1-1552) x Q14789 (1-3259)  
 QY 1 CTTCGGATGTCAGTCCGAGAAAGCTTAACCTTACTTAT-----GAG 39  
 DB 2204 VALLIENSPGLUALALysLysTTPGLUArgLysPheSerAspAlaIleGlnSerLysGlu 2223  
 QY 40 GAAAGAGTGTGAATTTGAGAAATCCCAATTGAAGTTTGTAGACAGACTTACGTGAATAT 99  
 DB 2224 GLUCIUleArgLysLysGlnAspAsnGlySerValIleLysAspGlnIleArgGlnMet 2243  
 QY 100 CAGAGAACTTGTGAAGACTTTAA-----GAGCACTAAGACAT----- 138  
 DB 2244 SerIleHisMetGlnGluLysIleAsnIleSerArgLysGlnIleAspLysGlnIle 2263  
 QY 139 -----AAGAAATTTCTGTGCTGTATACTTGTAAACCGTGTGGT 180  
 DB 2264 TTPGLuSerLysAlaGlnThrGlnValGlnIleGlnIleValCysAspThrLysGln 2283  
 QY 181 GGTCTTTGTTGAATGTGTCTCAGACATGAAAGCTTTCTTCCAAACCGATTAATGTT 240  
 DB 2284 GLY-----GluAsnLysGluLysSerGlnIleGluIleGluThrArg 2297  
 QY 241 CATATG-----CAGACCATCGAAAGACTGGTTAAAGAAAGATGACTTG 285  
 DB 2298 HisLeuThrHisSerSerGlnAsnGlnIleValLysLysGlnIleLysSerLeu 2317  
 QY 286 ATGTCTGACATGATTTCCGTGAAGAGCAAGCTTG----- 318  
 DB 2318 LysAspGlnIleThrAspLeuSerAsnSerIleGlnCysLysGlnIleLysGln 2337  
 QY 319 -----CSAGATGAGCAAGAAAGAAAGCAAGTGCCTAT 351  
 DB 2338 LeuGlnGlyIleIleArgGlnGlnIleAlaAspIleGlnAsnSerLysPheSer---Tyr 2356  
 QY 352 GAACAGGTGAACAAAGTTTGCAAATATCT-----GAGAA 387  
 DB 2357 GlnIleLysGlnIleThrAspLeuGlnIleSerArgLysGlnIleThrSerArgLysGlnIle 2376  
 QY 388 GCSAAATTTTGAAGAAACCAAGCGTTTA----- 414  
 DB 2377 IleAsnMetLysGlnGlnLysIleIleSerLeuLeuSerGlnLysGlnIleAlaIleGln 2396  
 QY 415 ATCCAGTGTGACAGTTGAGAGAGAGAGCTGAGAGGAGGAGCGAGCACTTGAAGAAAGAA 474  
 DB 2397 ValAlaIleAlaGlnIleArgGlnGlnIleHisAspLysGlnIleLysGlnIleGlnIle 2416  
 QY 475 CTTCATCTCAGCAAGAGAAAGGCGCATTCAGAAAGACATGATGAAGAAAGAAATTAAGC 534  
 DB 2417 LeuSerGlnGlnIleGlnIleGlnIleValIleGlnIleGlnIleLysAlaValAsp 2436  
 QY 535 AAAGAAAGGAGTACATCGGATCAAGATGTGTCTTCTGCAAGATATATGCCACTG 594  
 DB 2437 LysThrAsnGlnIleMetGlnIleThrLysThrIleLysLysGlnIleGlnIle 2455  
 QY 595 GAGCGCCAGGAGGAAAGGTTACAAAGAAAGATTCAGCTATTATAT----- 642  
 DB 2456 LysAlaGlnIleAspSerPheValLysSerMetSerIleGlnIleAsnAspArgAspArg 2475  
 QY 643 -----CAACTGGAGAA----- 654  
 DB 2476 IleValGlnAspArgLysGlnIleGlnIleArgHisLeuSerIleIleLysGlnLysAsp 2495  
 QY 655 -----ATTCAAACCCAGCTGCTTCGCGAAATGATGTCACAAAGGCTGTGGAGAA 708  
 DB 2496 GlnIleuIleGlnIleAlaIleAlaGlnAsnLysLysLysGlnIleIleArgGly--- 2514  
 QY 709 ATGCGCTATGACGTGAATTAACCAACATGAGAGAGATGAGCGAGAAAGGAGCAAGAGA 768  
 DB 709 ATGCGCTATGACGTGAATTAACCAACATGAGAGATGAGCGAGAAAGGAGCAAGAGA 768

```

Db 2515 LeuArgSerHisMetAspSerLeuAsnSerGluAsnAlaLysLeuAspAlaGluLeuIle 2534
QY 769 GAGTTCAGAGCAAAACAACTACAGGGATCTTGAAATTAAAGAT---CAGGAAATAGAGAA 825
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2535 GluTyrArgGlnAspLeuAsnGlnIleThrIleLysAspSerGlnGlnLysGlnLeu 2554
QY 826 TTGAGATAGAACTGGATGAGCAACAA---CAGTTGGAACAG 867
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2555 LeuGluValaGlnLeuGlnIleAsnLysGluLeuGluAsnLysTyrAlaLysLeuGlnIle 2574
QY 868 GAGCAGCAAAAGGACCCCTGGCCAGAGAGAGAGTGGCTGAGACTACAGAACTGCTGGCC 927
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2575 LysLeuLysGlnSerLeuGlnIleAsnGlnLysLeuArgSerPheAsnAlaLeuGln 2594
QY 928 GAATCTGAGACCAACTG---CAGCTC 951
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2595 GluGluLysGlnAspLeuSerLysGluIleGlnSerLeuLysValSerLieserGlnLeu 2614
QY 952 ACCAGATCTGAAATAGCTCACTCACTCAAGAA--- 984
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2615 ThrArg---GlnValThrAlaLeuGlnGlnGluGlyThrLeuGlyLeuTyrHisAlaGln 2633
QY 985 -----AAAGATATCATATGATATAATTGGGAAAGTTA-----CAGAGA 1023
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2634 LeuLysValaLysGlnGluGluValaHisArgLeuSerAlaLeuPheSerSerGlnLys 2653
QY 1024 AGAAATGAAAGATTGGAGGAGACAG---TGTGTCACAGCAT---CAGAGT 1059
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2654 ArgIleAlaGlnLeuGlnIleGlnLysValaLysValaGlnLysAlaLysLysVal 2673
QY 1060 GGGAGAGTATGAGACAGTGAAGCAAGGCTAAGCCAGCTGAT----- 1104
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2674 GlyIleIleGlnLysLeuLysLysGlnLysHisLeuHisAspAlaGlyLe 2693
QY 1105 ---AAGCAGAGCCAGCCAGCCAGCCAGCTGCTG---CAGCTCCTC 1146
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2694 MetAlaGlnGlnIleThrGlnThrAlaGlnGlnValaAlaGlnLeuAlaArgAspLeuVal 2713
QY 1147 AGCAAGCAGACAGCTTCTCTG-----GAGAGCAGAGAGCTGTGGAAGAGTG 1197
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2714 GluMetGlnGlnLysLeuLeuMetValThrLysGluAsnLysGlyLeuThrAlaGlnIle 2733
QY 1198 GACCGGCTCGGAGCCAGTACCAGCATGCCACATCT 1236
Db 2734 GlnSerPheGlyArgSerMetSerLeuGlnAsnSer 2746

RESULT 13
O9WVOO PRELIMINARY: PRT: 1022 AA.
AC O9WVOO:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Sperm tail associated protein.
CN STAP OR STAP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
OX NCB1_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Ohuchi J., Arai T., Kon Y., Watanabe T.;
RT "Sperm tail associated protein."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB029919; BAA82514.1; -.
KW MGP; MGI:1930136; Stap.
SQ SEQUENCE 1022 AA; 119400 MW; 4C41FE64AFCD984A CRC64;

```

```

Percent Similarity: 42.58% Conservative: 104
Best Local Similarity: 22.27% Mismatches: 158
Query Match: 9.22% Indels: 136
Db: 11 Gaps: 23

US-09-502-945-1 (1-1552) x O9WVOO (1-1022)
QY 10 GCATCCGAGAGCACTAAACTTATGAGAGAAAGCTGTGAATTCAGAAATCCCAATTC 69
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 174 AlaGlnGlnIleLeuAlaLeuAlaGlnLysPheLysLeuSerLeuArgSer---Leu 192
QY 70 AAGTTTTCAGACAGCACTTACCTGATATTCAGAGAACTTGTGAGACTTTAAAGGCAA 129
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 193 AsnLeuTyrArgAsp-----LysTyrGlnThrSerLeuSerAsnIle---GluLeu 208
QY 130 CTAAAGCATAAAGAAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 189
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 209 LeuGlnCysGlnValLysMetLeuGlnGlnLysLeuSerGlyLeuIleGlnAspPro 228
QY 190 TTGAATGTCCTCAGCATGAGCACTGCTTCTTCCCAACCCATATGATGTT----- 240
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 229 GluAsnLysGlyAspHisProLysValArgIleTyrThrSerProCysValIleGlnIle 248
QY 241 CATATCCAGACCACTGCAAGACTG-----GTTAAAGAAAGATGAC 282
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 249 HisGlnGlnIleThrLeuLysArgLeuSerGluValaTTPGlnLysValSerGlnGlnAsp 268
QY 283 TTGATG----- 288
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 265 LeuIleGlnGlnLeuArgAsnLysLeuAlaCysSerAsnSerLeuValLeuGlnArgIle 288
QY 289 TCTGCATGATTTCCCTGAAGAGCACTTGGCAGAT-----ACGAGCAAGAGAA 339
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 289 GluAlaLeuIleLysLeuGlnIleAlaGlnPheAlaSerTyrThrAlaThrHisArgHisPro 308
QY 340 GCAAGTGTATGACAGAGTGAACAAAGTTTGCAATTCAGAGAACCAATTTGAA 399
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 309 ProThrSerSerGlnAspCysGlnAspIleThrLysIleLeuLysHisLeuGlnGlnIle 328
QY 400 AAACCCAGAGCTTTAATCCAGTGTGACAG-----TTGAGAGAGAGCTGAGAG 450
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 329 LysAspSerGlnCysLeuHisValaGlnGlnIleTyrGlnAsnLeuValLysAspLeu 346
QY 451 CAGGCGAGCACTTGAAGAAAGACTTCATCTCAGCAAGAGAGAAAGCCATTGAGAA 510
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 347 -----ArgMetClnLeuGlnLysValaLysSerGlnGlnLysLys---IleMetLys 362
QY 511 GACATGATGAAAGGAATA----- 531
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 363 AspMetMetLysLeuGlnLeuAspLeuHisGlyLeuArgGlnIleThrSerCysValIle 382
QY 532 -----ACGAAAGAGAGAGTACATGAGATCAAGATGTTGATCTGTCTCAGAT 582
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 383 GluLysLysAspLysGlnThrValaPheLeuGlnIleTyrArgLeuGlnAspLeuGlnGln 402
QY 583 ATTGCCCAA-----CTGAGGCCCGAGGTGGAAGGTTTACAAAGCAAGAAATTTCA 633
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 403 TyrThrGlnSerGlnLysSerLeuLysLysAspLysLeuGlnIleAspLysAspGln 422
QY 634 GCTATTATCACTGAGAG-----GAAATTCAAAGCAGAGTGTCTCTGCGGAA 681
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 423 ArgLeuHisGlnLeuGlnLysAsnLeuMetGlnValaGlnAsnSerLeuArgGlnLysGln 442
QY 682 ATGATGTCACAAAGGTGTGTGAGAAATGCGCTATACACTGTAATTAACCAACATGAG 741
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 443 AlaGlnLeuGlnLys-----LeuGlnCysThrThrLysGln 454
QY 742 AAGATGAGCGCAAGAAAGAGCAGAGAGTTTCAGAGCAAAACTAATACAGGATCTTGA 801
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 455 LeuAspThrSerLeuGlnGlnLysArgGlnSerThrSerLysIle-----AspCysGln 472
QY 802 ATTAAAGATCAGGAATATGAGAAATTTGAGAAATGAACTGATGAAAGCAAGACCTTG 861
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Alignment Scores: 2.38e-08 Length: 1022  
Score: 252.00 Matches: 114



```

Db 473 AAlaenurAgaLagluileglnLysleuLysaspSerleuGlulAlaargGlulnleu 492
Oy 862 GAGACGAGAGCAGACAGAGCCCTGGCCAGAGAGAG----- 900
Db 493 LysValSeraspGlnasnLeuThrInSerLysGlulAlaHisLeuSerAlaSer 512
Oy 901 -----TGCCTG-----AGACTAACAGAA 918
Db 513 LeuGlnAspAlaHisArgLysIleGlulnscysLeuLeuGlnAspLysGlulnLysGluln 532
Oy 919 CTGCTGGCGCAATCTGACACCACTGCACTCACCAGATGTGAATATGCTCACTACAGT 978
Db 533 ValIleLysaspLeuGlnSerGlnLeuHisLysLeuGlnLysLysSerLysIleGlu 552
Oy 979 CAGAAAAAGGTATACATATGATTAATTTGGAAAGTTA----- 1017
Db 553 GlulGlnAlaGlyLysHisAsnArgGlnIleArgLeuGlnGluLeuSerSerGlnLeuSerGlu 572
Oy 1018 CAGAGAAAG-----AATGAAGAAATTGGAGAAACAGATGTCTCCAGCATGCGAGAGTACAT 1071
Db 573 GlnArgArgLeuSerAsnAlaGlnLysGlnLysSerLeuLeuGln----- 587
Oy 1072 GAGACGATGAAGCAAGGCTTAAGCAGCTGATTAAG-----CACAGCCAGGCCACACCC 1125
Db 588 LysThrLeuAspGlnAspGlnLysLysIleAspGlnLeuPheHisSerThrGlnValSer 607
Oy 1126 CAGCAG-----CTGGTGACGCTCCACACAGCAGCAACACCACTTCTCTCGAG 1173
Db 608 GlulnLysGlnArgGlnLeuThrAsnSerIleArgLysLeuGlnGlnLysLeuGln 627
Oy 1174 -----AGCGAG 1179
Db 628 IleLysGlnLeuLeuGlnGlnLysArgGlnGlnLeuLysLysSerLysGlnGlnLys 647
Oy 1180 AGCTGTGCGAAGAGGTGACCGCGCTGCGAGCCAG 1215
Db 648 AlaLeuGlnGlnLulileglnAlaLeuArgGlnGlu 659

RESULT 14
Oy 09BM8 PRELIMINARY; PRT; 876 AA.
ID 09BM8
AC 09BM8;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Paramyosin.
OS Sarcophaga scabiei.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcopitiformes; Astigmata; Sarcopitidae; Sarcopitidae;
OC Sarcophaga.
OC NCBI_TaxID=52283;
RN [1]
RP SEQUENCE FROM N.A.
RA Mattsson J.G., Ljunggren E.L., Bergstrom K.;
RT "Paramyosin from the parasitic mite Sarcophaga scabiei: cDNA cloning
RT and heterologous expression.";
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF317670; AA01181.1;
DR Pfam; PF01576; Myosin_tail.
DR InterPro; IPR002928; Myosin_tail.
SQ SEQUENCE 876 AA; 102454 MW; C99475EE7A0DA52 CRC64;

Alignment Scores:
Pred. No.: 2.77e-08 Length: 876
Score: 251.00 Matches: 109
Percent Similarity: 43.33% Conservative: 89
Best Local Similarity: 23.85% Mismatches: 173
Query Match: 9.19% Indels: 86
DB: 5 Gaps: 17

US-09-502-945-1 (1-1552) x 09BM8 (1-876)
Oy 16 GAGAGCTAAACTTACTTATGAGGAAAGTGTCAA-----ATTAGGAA 60

```

```

Db 374 GlulLysLeuAsnLeuAspLeuLysAsnLysLeuGlnGlnValThrMetLeuMetGluGln 393
Oy 61 TCCCAATGAAAGTTTGTGAGCAACACTTACGTGAATATCGAGAACTTGGAAGATCTT 120
Db 394 AlaGln---LysGlnAlaArgAlaLysAlaAlaGlnLeuGlnLysLeuGlnHisLulThr 412
Oy 121 AAAGCAACATTAAGCATTAAGAAATTTCTGCTGCTGCTGAATACTTGAACCGTTGTGT 180
Db 413 ---GluLysLeuAlaArgAspGlnArgAspAlaLeuAlaArgGlnAsnLysLysLeuThrAsp 431
Oy 181 GCTCTTTGTTGAAATGTGCTCAGCATGAGAGCTGTTCTTCCCAACCATTAATGTT 240
Db 432 AspLeu-----AlaGlnLysLysSerGlnLeuAsnAspAlaHisArgArgIle 447
Oy 241 CATATGCAGACCATTCGA-----AGACTGGTTAAAGAAAGATGACTGATGCTGCA 294
Db 448 HisGlnGlnGlnIleGlnIleLysArgLeuGlnAsnGlnArgGlnLysLeuSerAlaAla 467
Oy 295 CTAGTTCCGTAAGGAGCAGCTTGCCAGATACG-----CAGCAAGAGAAAGCAGTCT 348
Db 468 TyrLysGln-----AlaGlnThrLeuAlaGlyGlnGlnLulAlaAsn 482
Oy 349 TATGAACAGGTGAAACAAGTTTTCGAATATCTGAGAAAGCCAAATTTGAAAAACCAAG 408
Db 483 GlnArgLeuThrAlaGlnLysAlaGlnValArgHisAspTyrGlnLysArgLeuAlaGln 502
Oy 409 GCTTAATCCAGTGTGACCACTTGAGAAAGAGCTGAGAGAGCAGCGAGCACTTGA 468
Db 503 LysGlnGlnGlnIleGlnAlaLeuArgLysGlnTyrGlnIleGlnIleGlnLysLeuAsn 522
Oy 469 AAAGAACTTGATCTCAGCAAGCAAGAAAGGCGCATTTGAAGAAAGCATGATGAAGAA 525
Db 523 MetArgLeuAlaGlnAlaGlnLulLysLeuLysThrGlnLulAlaArgLeuLysLys 542
Oy 526 -----GAAATACGAAA-----GAAAGGAGTACATG 552
Db 543 TyrGlnAlaGlnIleThrGlnLeuGlnLeuSerLeuAspAlaAlaAsnLysAlaAsnIle 562
Oy 553 GGATCAAGAGATGTTGATCTGCTCAG-----AATATGCCCCAAGTGGAGCCAGGCTG 606
Db 563 AspLeuGlnLysThrIleLysLysGlnAlaLeuGlnIleThrGlnLeuGlnAlaHisThr 582
Oy 607 GAAAGGTTTACAAAGCAAGATTTACAGTATTAATCACTGAGCAAGAAATCAAGCCAG 666
Db 583 AspGlnValHisArgGlnLeuGlnGlnAlaValAspGlnLeuGlyValThrGlnArgArg 602
Oy 667 CTGCGCTCTCGGAATGATGATGCACAAAGGTGTGTGGAAGAAATGCGTATCAGCTGAAT 726
Db 603 CysGlnAlaLeuGlnAlaGlnLulLeuGln-----GlnGlnArgIleAlaLeuGln 618
Oy 727 AAACCAACATGAGAGAGATGAGCGCAAGAAAGCAGCAGAGACTTCAGACAAAACCT 786
Db 619 GlnAlaAsnArgAlaLysArgGlnAlaGlnLysLeuHisGlnGlnAlaValAlaArgVal 638
Oy 787 AAC----- 789
Db 639 AsnGlnLeuThrThrIleAsnValAsnLeuAlaSerAlaLysSerLysLeuGlnSerGlu 658
Oy 790 -----AGGATCTTGAATTAAGATCAG 813
Db 659 PheAlaAlaLeuGlnAsnAspTyrAspLysLysGlnLysGlnLeuArgIleSerThrAspLys 678
Oy 814 GAAATAGAGAAATTTGAGATTAAGACTGATGAAGCAACAAACACTTGAACAGAGCAG 873
Db 679 ArgValGlnLysLeuThrIleGlnLysSerThrLysAspLeuLeuValGlnGlnGln 698
Oy 874 CAGAGGAGAGCCCTGCGCAGAGAGAGGCGCTGACACATCAAGAAAGCTCTGGGCAATCT 933
Db 699 GlnArg-----LeuValLysMetGlnThrValLys-----LysSerLeuGlnGlnLul 714
Oy 934 GAGCACCATGACCTCACCAGATCTGAATAGCT-----CAACTGATCAAGAAAAA 987

```



```

QY 1051 ---GTCACGATGGAGAGTACATGAGACGATGAGCAAGCCTAAGCAGCTGATAG 1107
      ::::: ||| ::||| ||::: ||| :::
Db 639 GluLeuGluLeuArgAlaMetValGluGlnSerGluArgAlaArgLysLeuAlaGlu 658
      ::::: ||| ::||| ||::: ||| :::
QY 1108 CACAGCCAGGCCACAGCCACAGCTGTGTCAGCTCTCTACAGCAGAAAC---CAGCTT 1164
      ||| ::: ||||| ||||| ||||| |||
Db 659 GlnGluLeuIleGluAlaSerGluArgValGlnLeuLeuHisSerGlnAsnThrSerLeu 678
      ||| ::: ||||| ||||| ||||| |||
QY 1165 CTCCTGGAGAGCGCAGACCTGTGCGAAGAGGTGACCGGCTGCGGAGCCAGTTACCCAGC 1224
      ::: ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
Db 679 IleAsnGlnLysLysMetGluAlaAspIleSerGlnLeuGlnThrGluValGluGlu 698
      ::: ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
QY 1225 ATGCACACATCTGATTGC 1242
      ||| ::::: |||
Db 699 AlaIleGln---GluCys 703
      ||| ::::: |||

```

Search completed: March 21, 2003, 12:57:58  
 Job time : 81.2241 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 21, 2003, 11:21:37 ; Search time 14.6271 Seconds

(without alignments)  
8801.668 Million cell updates/sec

Title: US-09-502-945-1

Perfect score: 2732

Sequence: 1 cttctgagatcgcacgcagaaa.....aaatgaacttttaagaaga 1552

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool/US09502945/rnauc\_14032003\_101058\_19100/app\_query.fast\_1.10979  
-DB=SwissProt\_40 -QFMT=fastan -SUFFIX=isp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi -DIST=45  
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09502945\_6CGN\_1\_1\_113\_6runat\_14032003\_101058\_19100 -NCP=6 -ICPU=3  
-NO\_XLPRX -NO\_MMAP -LARGEOUTER -NEG\_SCORES=0 -WAIT -LONGLOG -DEV=TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	264	9.7	3911	AKA9_HUMAN	099996 h a-kinase
2	256.5	9.4	1790	USO1_YEAST	P25386 saccharomyc
3	255.5	9.4	978	RASO_AOAE	067124 aquifex aeo
4	255.5	9.4	1959	MYH9_CHICK	P14105 gallus galli
5	254	9.3	2017	MYSL_DROME	099323 drosophila
6	248.5	9.1	4687	PLEL_RAT	P30427 rattus norv
7	245	9.0	2245	MYSL_DICDI	P54697 dictyostel
8	244	8.9	1940	MYH3_CHICK	P02565 gallus galli
9	244	8.9	4684	PLEL_HUMAN	015149 homo sapien
10	244	8.9	1960	MYH9_HUMAN	P35579 homo sapien
11	242	8.9	2663	CENE_HUMAN	002224 homo sapien
12	241.5	8.8	1755	PEPL_MOUSE	091269 mus musculi
13	241.5	8.8	1938	MYSS_CHICK	P13538 gallus galli
14	240.5	8.8	1976	MYHA_HUMAN	P35580 homo sapien
15	239.5	8.8	880	MYSP_BRUMA	001202 bruma mala
16	239	8.7	879	MYSP_ONCVO	002171 onchocerca
17	238	8.7	848	MYSP_DIRIM	P13332 dirofilaria
18	237	8.7	1937	MYH8_HUMAN	P13335 homo sapien

19	236.5	8.7	1934	1	MYH7_MESAU	P13540 mesocricetu
20	236	8.6	1935	1	MYH7_RAT	P02564 rattus norv
21	236	8.6	1940	1	MYH3_RAT	P12847 rattus norv
22	235.5	8.6	1976	1	MYHA_BOVIN	Q27991 bos laurus
23	235	8.6	1130	1	YL17_CAEL	Q11102 caenorhabdi
24	235	8.6	1972	1	MYHB_RABIT	P35748 oryctolagus
25	234.5	8.6	1939	1	MYHA_HUMAN	095623 homo sapien
26	234	8.6	1549	1	TRHY_SHEEP	P22793 ovis aries
27	234	8.6	1756	1	PEPL_HUMAN	060437 homo sapien
28	234	8.6	1935	1	MYH7_PIG	P79293 sus scrofa
29	234	8.6	1961	1	MYH9_RAT	Q62812 rattus norv
30	234	8.6	4473	1	PLEL_CRIGR	Q91155 cricetus
31	233.5	8.5	794	1	HMHR_MOUSE	000547 mus musculi
32	233.5	8.5	1433	1	RESE_CHICK	Q42184 gallus galli
33	233.5	8.5	1818	1	HMW2_MYCPN	P75471 mycoplasma
34	233.5	8.5	1935	1	MYH7_HUMAN	P12883 homo sapien
35	233.5	8.5	1935	1	MYSS_CYPCA	Q90339 cyprinus ca
36	233.5	8.5	1938	1	MYA_AEOIR	P24733 aequipecten
37	232.5	8.5	976	1	SCPI_HUMAN	Q15431 homo sapien
38	232.5	8.5	1087	1	AKA9_RABIT	Q28658 oryctolagus
39	232.5	8.5	1939	1	MYH1_HUMAN	P12882 homo sapien
40	232.5	8.5	1957	1	YD66_SCHPO	Q10411 schizosacch
41	232.5	8.5	1976	1	MYH8_RAT	Q91140 rattus norv
42	232	8.5	1938	1	MYHD_HUMAN	Q9ukx3 homo sapien
43	232	8.5	1972	1	MYHB_MOUSE	Q08638 mus musculi
44	231.5	8.5	2704	1	BPA1_HUMAN	Q03001 homo sapien
45	231.5	8.5	3210	1	CENE_HUMAN	P49454 homo sapien

#### ALIGNMENTS

##### RESULT 1

AKA9\_HUMAN STANDARD; PRT; 3911 AA.  
ID Q99996; Q9U004; Q9U003; Q9Y6T2; Q14869; Q43355; Q94895; Q9Y6B8;  
AC Q99996; Q9U004; Q9U003; Q9Y6T2; Q14869; Q43355; Q94895; Q9Y6B8;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE A-kinase anchor protein 9 (Protein kinase A anchoring protein 9)  
DE (PRAK9) (A-kinase anchor protein 450 kDa) (AKAP 450) (A-kinase anchor  
protein 350 kDa) (AKAP 350) (hgAKAP 350) (AKAP 120 like protein)  
DE (Hyperion protein) (Yotiao protein) (Centrosome- and golgi-Localized  
PRN-associated protein) (CG-NAP).  
GN AKAP9 OR AKAP450 OR AKAP350 OR KIAA0803.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 4).  
RC TISSUE=Brain;  
RA Lin J.W., Wyszynski M., Madhavan R., Sealock R., Kim J.U., Sheng M.,  
Yotiao, a novel protein of neuromuscular junction and brain that  
interacts with specific splice variants of NMDA receptor subunit  
NRL.";  
RT J. Neurosci. 18:2017-2027(1998).  
RL [2]  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.  
RX MEDLINE=99219864; PubMed=10202149;  
RA Witczak O., Skalniak B.S., Keryer G., Bornens M., Tasken K.,  
Jahansen T., Oerstavik S.;  
RT "Cloning and characterization of a cDNA encoding an A-kinase anchoring  
protein located in the centrosome, AKAP450.";  
RL EMBO J. 18:1858-1868(1999).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 3).  
RC TISSUE=Brain;  
RA Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H., Ono Y.,  
RT "Characterization of a novel giant scaffolding protein, CG-NAP, that  
anchors multiple signaling enzymes to centrosome and the golgi  
apparatus.";



```

FT CONFLICT 3833 3833 T -> S (IN REF. 3).
SQ SEQUENCE 3911 AA; 453664 MW; 3FB1CB1C819B47AA CRC64;

Alignment Scores:
Pred. No.: 1.32e-07 Length: 3911
Score: 264.00 Matches: 121
Percent Similarity: 40.57% Conservative: 107
Best Local Similarity: 21.53% Mismatches: 162
Query Match: 9.66% Indels: 172
DB: 1 Gaps: 22

US-09-502-945-1 (1-1552) x AK9_HUMAN (1-3911)
OY 1 CTTGTGATGATCCGAGAGCTTAACCTTACTTATGAGAAAGGTGAA---ATTGAG 57
Db ::::::::::: ||| ::::::::::: ||| |||
Db 1855 MetLeuAnIleSerSerIleuGlnAlaIValGluLysLeuLeuGlnAlaIleSer 1874
OY 58 GAA-----TCCCAATTGAACTTTTGGAGAACGAC 87
Db ::::::::::: ||| ::::::::::: |||
Db 1875 GluThrSerSerGlnLeuGlnHisAlaLysValThrGlnThrGlnLeuMetIArgLys 1894
OY 88 TTAGTGCATATCAGAGAACTTGTAGACATCTTAAGAACCACTTAAGCATTAAGAAATT 147
Db ::::::::::: ||| ||| ||| ||| ||| ||| |||
Db 1895 PheArgGlnLysGlnGlnAlaThrGlnSerLeuLysCysGlnGlnLysValArg 1914
OY 148 CTTGTGGCTGCTAATCTGTAAACGCTGTGGTGGCTTTGTTGAATGTGTCAGCAT 207
Db ||| ::::::::::: ||| ||| :::::::::::
Db 1915 LeuHisGlnGlnLysSerArgAlaArgGlu-----GlnLeuAlaIValGlnLeuSerLysAla 1932
OY 208 GAAGCTGTTCTTCCCAAAACCCATATGCTTATGATGACAGCCATGAAAGACTGGTT 267
Db ||| ::::::::::: ||| ||| ||| ||| ||| |||
Db 1933 GluGlnValIleLeuPglTYTAlaAspGlnLysThrLeu-----PheGlnArgGlnIle 1950
OY 268 AAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 309
Db ::::::::::: ||| ||| ||| ||| ||| ||| |||
Db 1951 GlnGlnLysThrAspIleLeuAspArgLeuGlnGlnLysLeuLysAlaSerAsnArg 1970
OY 310 AGCACTTGGCAGATACGACAGCAAGA----- 336
Db ::::::::::: ||| ||| ||| ||| ||| |||
Db 1971 LeuGlnGlnLysGlnAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1990
OY 337 -----GAGCAAGTGGCTTATGAACAGGTAACCAAGTGGTTCGAA 375
Db ::::::::::: ||| ||| ||| ||| ||| |||
Db 1991 GlnLysGlnAlaMetLysAlaGlnAlaGlnLysProValGlnGln-----GlnLeuLeuGln 2008
OY 376 ATATGTGAGGAAGCAATTTGAAAGAACCAAGGCTTATATCCATGTCAGCAGTTCAGG 435
Db ::::::::::: ||| ||| ||| ||| ||| |||
Db 2009 GluThrGlnLysLeuMetLysGlnLysLeuGlnValGlnCysGlnAlaGlnLysValArg 2028
OY 436 AAGGAGCTGAGAGGCGGCGGAGCTTGAAGAAAGAACTGCAATCAG----- 486
Db ::::::::::: ||| ||| ||| ||| ||| |||
Db 2029 AspAspLeuGlnLysGlnValLysAlaLeuGlnIleAspValGlnGlnGlnValSerArg 2048
OY 487 -----CAAGAGAAAGGGCCATTGAGAAAGACATG----- 516
Db ::::::::::: ||| ||| ||| ||| ||| |||
Db 2049 PheIleGlnLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2068
OY 517 ---ATGAAAGAAAGAAATTAACGAAAGAGGATACATGGATCAAGATGTTG----- 567
Db ::::::::::: ||| ||| ||| ||| ||| |||
Db 2069 AlaLeuGlnLysGlnLeuGlnLysMetArgLysPheLeuAspGlnAlaIleLeuAspArg 2088
OY 568 -----ATCTGTCTCAGAAATATTGCCCAACTGGAGCCAGGTGGAGAAAG 612
Db ::::::::::: ||| ||| ||| ||| ||| |||
Db 2089 GlnHisGlnArgAspValPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2108
OY 613 GTTACAAAGAAAGAAATTTTACATTAATCACTGAGGAAATTTCAAGCCAGCTGGCT 672
Db ||| ::::::::::: ||| ||| ||| ||| ||| |||
Db 2109 ValProAlaArgPheGlnProIleSerGlnHisGlnThrArgGlnValGln-----GlnLeuAla 2127
OY 673 TCTCGGCAAAATGATGTCAAAAGGTGTGTGAGAAATG-----CGTATCAGCTG 723
Db ::::::::::: ||| ||| ||| ||| ||| |||
Db 2128 AsnHisLeuLysGlnLysThrAspLysCysSerGlnLeuLeuSerLysGlnGlnLeu 2147

```

```

OY 724 AATAAACAACATGAGAGAGATGAGCGAGAAAGACACAGAGATTCAGACAAAA 783
Db ::::::::::: ||| ||| ||| ||| ||| |||
Db 2148 GlnIArgAspIleGlnGlnIArgAsnGlu-----GlnIleGlnLysLeuGlnPheArgValArg 2166
OY 783 ----- 783
Db 2167 GlnLeuGlnGlnAlaLeuLeuValSerAlaAspThrPheGlnLysValGlnAspArgLys 2186
OY 783 ----- 783
Db 2187 HisPheGlnAlaValGlnAlaLysProGlnLeuSerLeuGlnValGlnLeuGlnAlaGln 2206
OY 784 -----ACTAAC----- 789
Db 2207 ArgAspAlaIleAspArgLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 2226
OY 790 -----AGGATCTTGAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGAT 843
Db ::::::::::: ||| ||| ||| ||| ||| |||
Db 2227 PheArgGlnGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 2246
OY 844 -----GAAAGCAACACACCTTGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 891
Db ::::::::::: ||| ||| ||| ||| ||| |||
Db 2247 IleGlnLysLysGlnLysThrThrArgLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2266
OY 892 AGAGAGAGAGTGGCTGAGACTACAGAACTGCTGGCGCAATGTGAG-----CAC 939
Db ::::::::::: ||| ||| ||| ||| ||| |||
Db 2267 LysAspAspMetGlnLysLeuGlnLysLeuAlaIleLysGlnSerAspAlaMetSerThrGln 2286
OY 940 CAACTGCACCTCCACAGATCTGAATAGCTCACTCACTCACTCACTCACTCACTCACTCACT 999
Db ::::::::::: ||| ||| ||| ||| ||| |||
Db 2287 AspGlnHisValLeuPheGlnLysPheAlaGlnIleIleGlnGlnLysGlnValGlnIle 2306
OY 1000 GATTAATTTGGA-----AAGTTACGAGAAAGA-----AAT 1029
Db ::::::::::: ||| ||| ||| ||| ||| |||
Db 2307 AspGlnLeuAsnGlnGlnValThrLysLeuGlnGlnGlnLysLysIleThrThrAspAsn 2326
OY 1030 GAAGAAATTGAGAGAA----- 1044
Db ::::::::::: ||| ||| ||| ||| ||| |||
Db 2327 LysValIleGlnGlnLysAsnGlnLeuLysLeuArgAspLeuGlnThrGlnIleGlnLysLeu 2346
OY 1045 -----CACTGTGTCAGCATGGAGAGATGATGATGATGATGATGATGATGATGATGAT 1092
Db ::::::::::: ||| ||| ||| ||| ||| |||
Db 2347 MetSerAspGlnGlnLysValLysArgAsnArg-----GlnGlnLysIle 2361
OY 1093 AGCGACGTGGATTAACACAGCCAGCCACAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 1152
Db ||| ||| ||| ||| ||| |||
Db 2362 GlnGlnLeuAsnGlnValIleGlnLysLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2381
OY 1153 CAGAACCAAGCTTCTCTGAGAGCGAGAGCCGTGCGAAGAGAGAGAGAGAGAGAGAGAGAG 1212
Db ::::::::::: ||| ||| ||| ||| ||| |||
Db 2382 ThrSer-----MetAsnAlaHisSerLeuSerGlnGlnAlaAspSerLeuLysHis 2398
OY 1213 CAGTTA 1218
Db ||| ||| ||| ||| ||| |||
Db 2399 GlnLeu 2400

RESULT 2
USOL_YEAST STANDARD: PRT, 1790 AA.
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Intracellular protein transport protein USOL.
GN USOL OR INT1 OR YDL058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
MEDLINE=91185402; PubMed=2010462;

```



RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,  
 RA Yamasaki M.;  
 RT "A cytoskeleton-related gene, usol, is required for intracellular  
 RT protein transport in *Saccharomyces cerevisiae*.";  
 RL J. Cell Biol. 113:245-260(1991).  
 RN [2]  
 RP SEQUENCE OF 782-1790 FROM N.A.  
 RA Hostetter M.K., Herman D.J., Bendel C.M., McCellan M., Tao N.,  
 RA Kendrick K.E.;  
 RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-8 FROM N.A.  
 RA Bai Y., Symington L.S.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI  
 CC COMPLEX.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR  
 CC MEMBRANS. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE  
 CC ER AND THE GOLGI COMPLEX.  
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED  
 CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL  
 CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.  
 CC -1- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: X54378; CAA38253.1; -;  
 DR EMBL: L03188; AAB00143.1; -;  
 DR EMBL: U53668; AAB66659.1; -;  
 DR PIR: A38455; A38455.  
 DR SGD: S0002216; USOL.  
 DR Interpro: IPR002017; Spectrin.  
 KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.  
 FT DOMAIN 1 724 GLOBULAR HEAD.  
 FT DOMAIN 725 1790 COILED COIL (POTENTIAL).  
 FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).  
 FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.  
 FT DOMAIN 1172 1786 ASF/GLU-RICH (ACIDIC).  
 FT CONFLICT 847 847 G -> E (IN REF. 2).  
 FT CONFLICT 924 924 E -> K (IN REF. 2).  
 FT CONFLICT 1253 1253 V -> I (IN REF. 2).  
 FT CONFLICT 1319 1319 I -> V (IN REF. 2).  
 FT CONFLICT 1461 1461 N -> S (IN REF. 2).  
 FT CONFLICT 1581 1581 G -> S (IN REF. 2).  
 FT CONFLICT 1600 1600 I -> V (IN REF. 2).  
 FT CONFLICT 1661 1661 R -> S (IN REF. 2).  
 FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).  
 SO SEQUENCE 1790 AA: 206424 MW: 6CE2B216E9FD4818 CMC64;  
 Alignment Scores:  
 Pred. No.: 3.72e-07 Length: 1790  
 Score: 256.50 Matches: 86  
 Percent Similarity: 43.33% Conservative: 99  
 Best Local Similarity: 20.14% Mismatches: 169  
 Query Match: 9.39% Indels: 73  
 DB: 1 Gaps: 11  
 US-09-502-945-1 (1-1552) x USOL\_YEAST (1-1790)  
 QY 64 CAATGAGTTTGGAGGAGCTTACGATATCAGAGAACTTGTGAATCTT--- 120  
 Db 963 LysLeuLysSerLeuAlaAsnAsnTrpLysAspMetGlnLaGluAsnGlnSerLeuIle 962  
 QY 121 -----AAGAGCACTAAAGCAATAAGATTTCTTGCTGCTCTAATATCTTAACCGT 174  
 Db 983 LysAlaValGlnGluLysSerLysAsnGlnSerSerIleGlnLeuSerAsnLeuGlnAsnLys 1002

QY 175 GTTGGTGGTCTTTGTTGAATGTGCTCAGCATGAGAGCTGTTCTTCCCAACCCACTACT 234  
 Db 1003 IleAspSer-----MetSerGlnGluLysGlu 1011  
 QY 235 AATGTTATATGAC-----ACATCGAAGACGTGTTAAAGAAAGA 276  
 Db 1012 AsnheGlnIleGlnArgGlySerIleGlnLysAsnIleGlnLeuLysTrpIle 1031  
 QY 277 GATGACTGTGATGTGACATAGTTCCGTAAAGACAGCTGGCAGATPACGACCAANAAGA 336  
 Db 1032 SerAspLeuGlnGlnThrLysGlnGlnIleLysSerLysSerAspSerLysAspGlu 1051  
 QY 337 GAACCAAGTGTATGAAACAGGTGAACAAAGTTTGCAAATATCTGAGGAAGCCAAATTT 396  
 Db 1052 TyrLysGlnGlnIleSerLeuLeuLysGlnLysLeuGlnThrAlaThrTrpAlaAsnAsp 1071  
 QY 397 GAAAAACCAAGGCTTTAATCCAGTGTGACCAAGTGGAGAAAGAGCTGGAGAGCAG--- 453  
 Db 1072 GluAsnValAsnLysIleSerGlnLeuThrLysThrArgGlnGlnGlnAlaGlnLeu 1091  
 QY 454 -----GGCAGAGCACTT 465  
 Db 1092 AlaAlaTrpLysAsnLeuLysAsnGlnLeuGlnThrLysLeuGlnThrSerGlnLysAla 1111  
 QY 466 GAAAGAGAACTTGATGATCTCAGCAGAGAGAAAGGCGCATTGAGAAAGACATGATGAAAG 525  
 Db 1112 LeuLysGlnValLysGlnAsnGlnGlnIleLysLysGlnGlnLysIleGlnLeuGlnLys 1131  
 QY 526 GAATATACGAAGAAGAGGAGCTACATGGCATCAAGATGTTGATCTGTCTCAAAATATTT 585  
 Db 1132 GluAlaThrGlnThrLysGlnGlnLeuAsnSer----- 1142  
 QY 586 GCCCAACTGGAGCCCGAGTGGAAAGATTCACAAAGAGAAAGATTTCAGCTATTAAATCAA 645  
 Db 1143 -----LeuAlaTrpAlaAsnLeuGlnSerLeuGlnLysGlnLysGlnLysAlaAlaGln 1160  
 QY 646 CTGGAGGAAATTCAGAACCCAGCTGGCTTCTCGGGAAATGATGTCACAAAGCTGTGGA 705  
 Db 1161 LeuLysLysTrpArgGlnGlnGlnIleAlaAsnLysGlnArgGlnTrpAsnGlnGlnLysLeu 1180  
 QY 706 GAATGGCGATATCGATCGAATTAACCAACATGAG-----AAG 744  
 Db 1181 GlnLeuAsnAspGlnIleThrSerThrGlnGlnGlnLysGlnLysLysLysAsn 1200  
 QY 745 GATGAGGCGAGAAAGGAGCAGACAGAGTTCAGCAAACTAAGAGGATCTTGAATTT 804  
 Db 1201 AspGlnLeuGlnGlnGlnValLysAlaMetLysSerThrSerGlnGlnLysAsnLeu 1220  
 QY 805 AAGATCAGCAATATAGCAAAATTGCAATATGAACTGATGAAAGCAACACTTGGA 864  
 Db 1221 LysLysSerGlnIleAlaSerAlaLeuAsnLeuGlnIleLysGlnLeuLysLys-----Lys 1238  
 QY 865 CAGAGCAGAGCAAGAGAGCGCTG-----GCCAGAGAGAGAGTGCCTG 906  
 Db 1239 AsnGlnThrAsnGlnThrAlaSerLeuLeuGlnSerIleLysSerValGlnSerGlnThrVal 1258  
 QY 907 AGACTTAACAGAACTGCTGGGCGAATCTGAGCACCACACTGACCTCAGACATCTGAATA 966  
 Db 1259 LysIleLysGlnLeu-----GlnAspGlnLysAsnPhelLysGlnLysGlnVal 1274  
 QY 967 GCTCACTCAGTCAAGAAAAAGGTATACATATGATATATTTGGGAAG-----TTA 1017  
 Db 1275 SerGlnLeuGlnLysLysLeuLysLysAlaSerGlnAspLysAsnSerLysTrpLeuGlnLeu 1294  
 QY 1018 CAGAGAAAGAAATGAGAAATTTGAGGAACAGTGTCCAGCATGGAGAGATACATGAGAGC 1077  
 Db 1295 GlnLysGlnSerLysLysIleLysGlnGlnLeuAspAlaLysThrArgLeuLysIle 1314  
 QY 1078 ATGAGCAAGAGGCTTAAGGAGCATTAAGCAGACAGCCAGCCAGCAGCTGCTG 1137  
 Db 1315 GlnLeuGlnLysLysLeuThrAsnLeuSerLysAlaLysGlnLysSerGlnLysLeuSer 1334  
 QY 1138 CAGCTCCTCAGCAAGAGCAAGCAAGCTTCTCTGAGAGAGCAGAGCTGTGGAAGAGGTG 1197

```

Db 1335 ArgLeuLysThrSerSerLnu-----GluArgLysAsnAlaGluGluGlnLeu 1351
QY 1198 GACCGGCTGGCGACCCACTTA 1218
Db 1352 GluLysLeuLysAsnGluLeu 1358

RESULT 3
RA50_AOUAE STANDARD: PRT: 978 AA.
ID RA50_AOUAE
AC 067124
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable DNA double-strand break repair rad50 ATPase.
GN RAD50 OR AO_1006.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
CC -1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -1- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC DR EMBL; AE000718; AAC07092.1; -
CC DR InterPro: IPR003439; ABC_transportr.
CC DR InterPro: IPR004592; SPC.
CC DR InterPro: IPR002017; Spectrin.
CC DR TIGRfams; TIGR00618; spcc.1.
CC KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
CC FT NP_BIND 32 39 ATP (By similarity).
CC FT DOMAIN 160 826 COILED COIL (POTENTIAL).
CC SEQUENCE 978 AA; 115897 MW; 9B0F2BF51ADD151 CRC64;
SQ

Alignment Scores:
Pred. No: 4.57e-07 Length: 978
Score: 255.50 Matches: 116
Percent Similarity: 42.16% Conservative: 107
Best local Similarity: 21.93% Mismatches: 157
Query Match: 9.35% Indels: 149
DB: 1 Gaps: 23

US-09-502-945-1 (1-1552) x RA50_AOUAE (1-978)
QY 16 GAGAGCTAAACTTACTATGAG-----GAAAGGTGAAGT 54
Db 185 GUAUAlaLeuLysGluTyrGluLeuLeuLysAspTyrThrProThrLysLysGluVal 204
QY 55 GAGGAATCCCAATGGAAGCTTTTGAAGACGACTTACGTGATATACAGAGAAGCTTGTA 114

```

```

Db 205 LeuGluLysThrLeuLysAsnLeuGluGluGlnLeuLysGluLeuLysGluThrGluGlu 224
QY 115 GATCTTAAGGACACACTAAGACATTAAGATTTCTTCGGCTGCTAATACTTGTAACGT 174
Db 225 LysLeuArgGlnGluLeuLysLysAlaGlu 234
QY 175 GTTGGTGGCTCTTGTGTAATGTCCTCAGCATGAGCTGTTCTTCCCAACCCACT 234
Db 235 -----GluLysAspSerLeuGluArgGluLeuSerGluValThr 248
QY 235 AATGTTCAATGACAGACCATGCAAGACTGTTAAGAA----- 273
Db 249 Lys-----LeuLysGluLeuGluAsnLeuGluLysGluValGluLysLeuArgGluLys 266
QY 273 ----- 273
Db 267 LeuGluPheSerArgLysValAlaProTyrValProIleAlaLysArgIleGluGluLeu 286
QY 273 ----- 273
Db 267 AspLysLysLeuThrGluLeuLysValArgLysAsnLysLeuThrLysGluLeuAlaVal 306
QY 274 ---AGAGATGACTTG-----ATGTCGCACTAGTTCCGTAAGAGC 312
Db 307 LeuLysAspGluLeuSerPheAlaGlnGluGluLeuAsnArgIleGluAlaGluLysGlu 326
QY 313 AGCTTGGCAGATACCCACCAAGAAAGAA-----GCAAGTGCTTATGAA 354
Db 327 LysPheLysGluGluLysGluArgGluLysGluLeuGluHisArgLeuLysLysLeuGln 346
QY 355 CAGGTGAACAAGTTTCTT---CAATATCTGAGGAAGCAATTTGAAAACCAAGGCT 411
Db 347 GluLysGluLysLeuLysGluLeuSerGlnLeuSerSerLeuLysGluLysGlu 366
QY 412 TTAATCCAGTGTGACCACTTGACGAGAGAGCTGCGAGCGGCGGACGACTTGA--- 468
Db 367 Arg---GluTyrGluGlnAlaLysGlnLeuPheGluAspLeuSerGluArgValGluLys 385
QY 469 ---AAGAACCTTGCACTTCACAGCAAGAAAGGCGCAATTTGAAACATGATGAAGAG 525
Db 386 GlyLysLysLeuValAlaGluThrGluGluLys---LeuGluLys-----IleLys 401
QY 526 GAAATTAAGAAAGAGGAGTACATGCGATCAACAGATG-----TTGATC 570
Db 402 GluLeuPheSerGluGluGluTyrThrSerLeuLysMetLysGluArgLeuLeuValGlu 421
QY 571 TTGCTCAGAAATATGCCCACTG-----GAGGCCAGGTGGAAGGTTACA--- 618
Db 422 LeuGlnArgLysLeuLysGluLeuLysGluLysGluGlnLeuGluAsnLeuThrGln 441
QY 619 -----AAGGAAGAATG-----TCAGTATTAATCAACTGGAGGAATTCGA 660
Db 442 LysTyrLysGluLysLysValHisGluLysValLeuAsnGluLeuLysGluLeuGlu 461
QY 661 AGCCGCTGGCTTCGGGGAATG----- 684
Db 462 ArgGluLeuLysGluArgGluLeuHisTyrHisAlaHisMetValAlaSerTyrLeuSer 481
QY 685 -----GATGTCACAAAGGTGTGTGAGAAATGCGTATCAG-----CTGAATAA 729
Db 482 ProGluAspThrProValLysGlyLysIle---TyrArgGlyLysAlaLeuAsn 500
QY 730 ACCAAGATGAGGAAGATGAGGACGAGCAAAAGGACACAGAGTTCAAGCAAAAACCTAAC 789
Db 501 ValAspAlaGluGlyIleSerGluLeuLysHisAlaLysGluLeuLysGluLysGluGlu 520
QY 790 AGGATCTTGAAGTAATAA-----GATCAGGAA 816
Db 521 ArgGluIleAspThrThrLeuLysLeuTyrAlaGlnLysIleAsnSerLeuLysGluGlu 540
QY 817 ATGAGGAATTTGAGATGAATGAGTGAAGCAACAAACACTTGGAACAGAGCAGCAGCAG 876
Db 541 MetGluLysLeuArgAsnGluValGluGluLeuLeuArgLysGluIle---ProGluAsnLeu 559

```



```

Db 1207 Gln-----ThrLysArgValLysAlaAsnLeuGluLys-----AlaLysGlnIle 1221
OY 529 ATACGAAAGAAAGGAGTACATGGATCAAGATGTTGATCTTCTCAGAAATATTTC 588
Db 1222 LeuGlnSerGluArgAlaGlnLeuSerAsnGlnValLysValLeuGlnGlyLysGly 1241
OY 589 CAACGTGAGGCCACGCTGGAGAAAGTTACAAAGAAATTCAGTATTAATCAACTG 648
Db 1242 AspaLagLuuHsLys-----ArgLysLysValAspAla-----GlnLeu 1254
OY 649 GAGGAAATTCAAAGCCAGCTGCTTCTCGGAAATGATGTCACAAAGCTGTGTGAGAA 708
Db 1255 GlnGluLeuGlnValLysPheThrGluGluLys-----ArgValLysThrGlu 1270
OY 709 ATGCGCTATCAGCTGATTAACCAACATGGAG----- 741
Db 1271 LeuAlaGluArgValAlaSnLysLeuGlnValGlnLeuAspAsnValThrGlyLeuLeuAsn 1290
OY 741 ----- 741
Db 1291 GlnSerAspSerLysSerIleLysLeuAlaLysAspPheSerAlaLeuGlnSerGlnLeu 1310
OY 742 AAGCATGAGCGAAGAAAGAGACAGAGAGTTCAGAGCAAAACAGAGGATCTTGA 801
Db 1311 GlnAspThrGlnGlnLeuLeuGlnGlnGlnLurThrArgLeuLysLeuSerPheSerThrLys 1330
OY 802 ATTAAGATCAGCAAAATAGACAATTTGAGATA-----GAACCTGGATGAA 846
Db 1331 LeuLysGlnThrGlnLysAspGlnLysAsnAlaLeuLysGlnGlnGlnGlnGlnGln 1350
OY 847 AGCAAAACACACTGGAGACAGAG-----CAGCAAGAGCGACCTGGCCGCA 894
Db 1351 AlaLysArgAsnLeuGlnLysGlnLysLeuSerValLeuGlnGlnGlnAlaValGlnLys 1370
OY 895 GAGGAG-----TGCTGAGACTAACAGAACTGCTGGCGAATCTGAG 936
Db 1371 LysLysMetAspAspGlyLeuGlyCysLeuGlnLysLeuAlaGlu-----GlnAlaLys 1387
OY 937 CACCAACTGCACCTCCACCATGATCTGAATAGCTCACTCAAGAAAAGGTATACA 966
Db 1388 LysLysLeuGlnLysAspLeuGlnSerLeuThrGlnArgTyrGlnGlnLysIleAla 1407
OY 997 TATGTAATTTGGGAAATCTTACAGAGAGA-----AATGAAGATGGAGAACAGTCTC 1053
Db 1408 TyrAspLysLeuGlnLysThrLysThrArgLeuGlnGlnLysLeuAspIleAlaVal 1427
OY 1054 CAGCATGGAGAGTACATGAGACGATGAAGCAAGGCTAAGCAGCTGATTAAGCAGC 1113
Db 1428 -----AspLeuAsp-----HisGln 1432
OY 1114 CAGGCCACAGCCACAGACCTGCTGACCTCTCAGACAGCAAGAACCTTCTCTCTGAG 1173
Db 1433 ArgGlnThrValSerAsnLeuGlnLysLysGlnLysLysPheAspGlnLeuLeuAlaGlu 1452
OY 1174 AGGAGAGCTCTGCTG-----GAGAGAGTGGACGGCTGGCGGAGCCAG 1215
Db 1453 GlnLysAsnIleSerAlaLysTyrAlaGlnLysArgAspArgAlaGlnAlaGlu 1470

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=90349606; PubMed=2117279;
RA Ketchum A.S., Stewart C.T., Stewart M., Kleihart D.P.;
RT Complete sequence of the Drosophila nonmuscle myosin heavy-chain
RT transcript: conserved sequences in the myosin tail and differential
RT splicing in the 5' untranslated sequence."
RL Proc. Natl. Acad. Sci. U.S.A. 87:6316-6320(1990).
CC -1- FUNCTION: NONMUSCLE MYOSIN APPEARS TO BE RESPONSIBLE FOR
CC CELLULARIZATION. REQUIRED FOR MORPHOGENESIS AND CYTOKINESIS.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM: ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M35012; AAA28713.1; -.
CC PIR: A36014; A36014.
CC PIR: B36014; B36014.
CC HSP: P10587; IBR2.
CC FlyBase: FBgn0005634; zip.
CC InterPro: IPR000048; IQ_region.
CC InterPro: IPR004009; Myosin_N.
CC InterPro: IPR002928; Myosin_tail.
CC InterPro: IPR001609; myosin_head.
CC Pfam: PF00612; IQ; 1.
CC Pfam: PF01576; Myosin_tail; 1.
CC Pfam: PF02736; Myosin_N; 1.
CC PRINTS: PR00193; MYOSINHEAVY.
CC ProDom: PD000355; myosin_head; 1.
CC SMART: SM00015; IQ; 1.
CC SMART: SM00242; Myosin_N; 1.
CC PROSITE: PS50096; IQ; 1.
CC Myosin: Alternative splicing: Colled coil; Actin-binding;
CC ATP-binding; Calmodulin-binding;
CC MYOSIN HEAD-LIKE.
CC DOMAIN 1 829
CC FT 830 859
CC FT DOMAIN 886 2017
CC FT NP_BIND 225 232
CC FT DOMAIN 250 260
CC FT DOMAIN 682 694
CC FT DOMAIN 705 727
CC FT DOMAIN 742 758
CC FT DOMAIN 1303 2017
CC FT DOMAIN 1303 1970
CC FT DOMAIN 1971 2017
CC FT VARSPLIC 1 45
CC FT SEQUENCE 2017 AA; 232016 MW; 73E3CB02BA8F2528 CRC64;

```

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	4.98e-07	254.00	110	101	174	104
Percent Similarity:		43.15%				
Best Local Similarity:		22.45%				
Query Match:		9.30%				

US-09-502-945-1 (1-1552) x MYSN\_DROME (1-2017)

```

OY 4 CTGATGATCCGAGAACTTAACCTTATATAGAGAAAGTGTGAATTTGAGGAATCC 63
Db 1143 IleAspGlnGlnSerAlaThrLysAlaGlnLysAlaGlnLysAlaGlnLysArgGlnLys 1162
OY 64 CAATTGAAGTTTGGAGAACGACTTA-----GCTGAATATCAG 102

```

```

Db 1163 GlnLeuLacIuIleGInGluAspLeuGluAgluLysAlaAlaArgAlaLysAlaGlu 1182
QY 103 AGAAGTCGTGAAGCTTTAAAGACCACTAAAG-----CATAAAGATTCTTCGCT 156
Db 1183 LysValAlaArgAspLeuSerGlnuLeuGluAlaLeuLysAsnGluLeuLeuAspSer 1202
QY 157 GCTAACTACTGTAAACCGCTGTGGTGTCTTTGTAATGTGCTCAG-----204
Db 1203 LeuAspThrAlaAlaGlnGlnGluLeuArgSerLysArgGlnGlnuLeuAlaThr 1222
QY 205 -----CATGACGTGTCTT-----219
Db 1223 LeuLysLysSerLeuGlnGluThrValAsnHisGlnuLysValLeuAlaAspMetArg 1242
QY 220 -----TCCCAACCCACTACTATGTCATATGCAGACACTGGAAGACTGTTAA 270
Db 1243 HisLysHisSerGlnuLeuLeuAsnSerIleAsnAspGln-----LeuGluAsnLeuArgLys 1261
QY 271 -----GAAAGAGATGACTGTATG 288
Db 1262 AlaLysThrValLeuGlnuLysAlaLysGlyThrLeuGlnuLysAlaAspLeuAla 1281
QY 289 TCTGCACTACTTCCCTAAAGAGACGCTTGCGACATACGACGAAAGAGCAAGTGT 348
Db 1282 ThrGluLeuArgSerValAsnSerSerArgGlnGlnuAsnAspArgArgLysGlnAla 1301
QY 349 TATGACACAGCTGGAACAAAGTTTGCACAAATCTGAGAGAACCAATTTGAAAAACCAAG 408
Db 1302 GlnSerGlnIleAlaGlu-----LeuGlnVal-----LysLeuAlaGluIleGluAlaArg 1319
QY 409 GCTTTAATCCAG-----TGTACCACTTGAGAGAGCTGAGAGGACGCGGAGCA 462
Db 1320 SerGlnuLeuGlnuLysCysThrLysLeuGlnGlnuLysAlaAsnIleThrAsnGln 1339
QY 463 CTGGAAGAA-----GAACTTGATCTGACCAAGAGAA 495
Db 1340 LeuGlnGlnuLysAlaSerAlaAlaValLysSerIleAsnMetGlnSer 1359
QY 496 AGGCACTTGAGAAAGACATGATGAAAGAAATACGAAAGAAAGAGATCATGCGA 555
Db 1360 GlnLeuThrGlnuLysGlnuLeuGlnuLysGlnuLysArgGlnuLys-----1375
QY 556 TCAAGAGTGTGATCTGTCTCAGAAATATGCCCACTGAGGCCCGAGTGAAGAGTT 615
Db 1376 -----LeuGlnLysSerLysLeuArgGlnIleGlnuSerGlnuLysGlnuLys 1392
QY 616 ACAAGAGAAAGATTCACGTATTATACATGAGAGAAATTCAAAGCCAGTGGCTCT 675
Db 1393 GlnGlu-----GlnLeuGlnuLysAspGlnuLysAlaLysArgAsn 1405
QY 676 CGGGAATGATGTCCAAAGGTGTGTGAGAAATGCGTATGACGTATTAACCAAC 735
Db 1406 TyrGlnuArgLysLeuAlaGlnuLysThrThrGlnMetGln-----GluIleLysLysLysAla 1424
QY 736 ATGAGAGAGATGAGCAGCAAGAAAGACAGACAGAGTTGACAGCAAAACTAACAGGAT 795
Db 1425 GlnGluAspAlaAspLeuAlaLysGlnuLysGlnuLysLysArgLeuAsnAsp 1444
QY 796 CTGGAATTAAGATCAGAAATAGAGAAATTCAGAAATAGAA-----CTGGATGAA 846
Db 1445 IleGlnAlaLeuGlnuLysGlnuLysGlnuLysLeuIleAlaGlnAsnAspArgLeuAspLys 1464
QY 847 AGCAACAACACTTGGACAGCAGACAGACAGCAGCC-----885
Db 1465 SerLysLysLysIleGlnSerGlnuLeuGlnuLysAlaThrIleGlnuLysGlnuLysArg 1484
QY 886 -----CTGGCCAGAGAG 897
Db 1485 ThrLysValLeuGlnuLysGlnuLysGlnuLysAsnPheAspLysLysIleLeuAlaGlnu 1504
QY 898 GAGTGGCTG-----AGACTAACAGAACTGCTGGGGAATCTGACCACTGACCTC 951

```

```

Db 1505 LysAlaIleSerGlnuGlnuIleAlaGlnuLysArgAspThrAlaGlnuAlaArgGlu 1524
QY 952 ACCGATCTGAATATGCTCACTACGTCAAGAAAGATATACATGATTAATGGA 1011
Db 1525 LysGlnuThrLysValLeuSerValSerArgGlnuLeuAspGlnuAlaPheAspLysIleGlu 1544
QY 1012 AACTTACAGCAAGAAATGCAAAATTCGAGCAACAGTGTCTCCAGCATGGAGAGTACAT 1071
Db 1545 AspLeuLysnLysArgLysThrLeuGlnuAsnGlnuLeuAspAspLeuAlaAsnThrGln 1564
QY 1072 GAGACGATGAGCAAGAGCTTAAGCAGCAGTGTATAGCAGACCGCAGACCCAGCAG 1131
Db 1565 GlyThrAlaAspLysAsnValHisGlnuLeuGlnuLysAlaLysArgAlaLeuGlnuSerGln 1584
QY 1132 CTGGTGCAGCTCTCCAGCAACAGCAACACCTTCTCTGAGAGAGCAGACCTGTGCGAA 1191
Db 1585 LeuAlaGlu-----LeuLysAlaGlnuAsnGlnuLysLeuLysAspLeuGlnuLeuThrGln 1603
QY 1192 GAGTGGACCGCTGCGGACCCAGTTA 1218
Db 1604 AspAla-----LysLeuArgLeuGlnuVal 1611

RESULT 6
PDEL_RAT
ID PDEL_RAT STANDARD; PRT; 4687 AA.
AC P30427; 008879; 008880; 008881;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Plectin 1 (PLTN) (PCN).
GN PLECL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Glial tumor;
RX MEDLINE=91268156; PubMed=2050743;
RA Wiche G., Becker B., Lubet R., Weitzer G., Castanon M.J.,
RT Hauptmann R., Stratowa C., Stewart M.;
RT "Cloning and sequencing of rat plectin indicates a 466-kD polypeptide
RT chain with a three-domain structure based on a central alpha-helical
RT coiled coil."
RL J. Cell Biol. 114:83-99(1991).
RN [2]
RP REVISIONS.
RC TISSUE=Glial tumor;
RX MEDLINE=96210632; PubMed=8633053;
RA Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;
RT "Human plectin: organization of the gene, sequence analysis, and
RT chromosome localization (8q24).";
RL Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).
RN [3]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4), AND TISSUE SPECIFICITY.
RC TISSUE=Glial tumor;
RX MEDLINE=97321050; PubMed=9177781;
RA Elliott C.E., Becker B., Oehler S., Castanon M.J., Hauptmann R.,
RT Wiche G.;
RT "Plectin transcript diversity: identification and tissue distribution
RT of variants with distinct first coding exons and rodless isoforms.";
RL Genomics 42:115-125(1997).
CC -1- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND
CC MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSOMES OR
CC HEMIDESMOSOMES. MAY BE INVOLVED NOT ONLY IN THE CROSSLINKING AND
CC STABILIZATION OF CYTOSKELETAL INTERMEDIATE FILAMENTS NETWORK, BUT
CC ALSO IN THE REGULATION OF THEIR DYNAMICS.
CC -1- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS: 1 (SHOWN HERE), 2, 3 AND 4; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSION IN
CC SKELETAL MUSCLE AND LOWEST IN THYMUS.
CC -1- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH

```

CC VIMENTIN, DESMIN, GFAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-  
 CC AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.  
 CC -1- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM  
 CC INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 1 CALPOTIN-BINDING DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 CALPOTIN-HOMOLOGY (CH) DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 3 PLECTIN REPEATS.  
 CC -1- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.  
 CC -1- SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.  
 CC -1- CAUTION: ISOFORM 4 IS A FRAGMENT AT THE N-TERMINUS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X59601; CAA42169.1; -  
 CC EMBL: U96274; AAC53209.1; -  
 CC EMBL: U96275; AAC53210.1; -  
 CC EMBL: U96276; AAC53211.1; -  
 CC PIR: A39638; A39638.  
 CC PIR: S21876; S21876.  
 CC HSSP: Q01082; 1BKR.  
 CC InterPro: IPR001589; Actbind\_actin.  
 CC InterPro: IPR001715; Calponin-like.  
 CC InterPro: IPR001101; Plectin\_repeat.  
 CC InterPro: IPR005326; S10\_plectin.  
 CC InterPro: IPR002017; Spectrin.  
 CC Pfam: PF00307; CH\_2.  
 CC Pfam: PF00681; Plectin\_21.  
 CC Pfam: PF03501; S10\_plectin\_1.  
 CC SMART: SM00033; CH\_2.  
 CC SMART: SM00250; PLEC\_33.  
 CC SMART: SM00150; SPEC\_4.  
 CC PROSITE: PS00019; ACTININ\_1; FALSE\_NEG.  
 CC PROSITE: PS00020; ACTININ\_2; FALSE\_NEG.  
 CC PROSITE: PS50021; CH\_2.  
 CC Colled coll; Repeat; Structural protein; Cytoskeleton; Actin-binding;  
 KW Phosphorylation; Alternative splicing.  
 KM DOMAIN 1 1473 GLOBULAR 1.  
 FT DOMAIN 1 1474 2758 CENTRAL FIBROUS ROD DOMAIN.  
 FT DOMAIN 2 2759 4687 GLOBULAR 2.  
 FT DOMAIN 181 406 ACTIN-BINDING.  
 FT DOMAIN 185 288 CH 1.  
 FT DOMAIN 301 403 CH 2.  
 FT REPEAT 648 722 SPECTRIN 1.  
 FT REPEAT 743 827 SPECTRIN 2.  
 FT REPEAT 840 933 SPECTRIN 3.  
 FT REPEAT 1318 1418 SPECTRIN 4.  
 FT DOMAIN 1472 1692 COILED COIL (POTENTIAL).  
 FT DOMAIN 1724 2760 COILED COIL (POTENTIAL).  
 FT REPEAT 2791 2828 PLECTIN 1.  
 FT REPEAT 2829 2866 PLECTIN 2.  
 FT REPEAT 2867 2904 PLECTIN 3.  
 FT REPEAT 2905 2942 PLECTIN 4.  
 FT REPEAT 2943 2980 PLECTIN 5.  
 FT REPEAT 2984 3018 PLECTIN 6.  
 FT REPEAT 3119 3156 PLECTIN 7.  
 FT REPEAT 3157 3194 PLECTIN 8.  
 FT REPEAT 3195 3232 PLECTIN 9.  
 FT REPEAT 3233 3270 PLECTIN 10.  
 FT REPEAT 3271 3308 PLECTIN 11.  
 FT REPEAT 3311 3346 PLECTIN 12.  
 FT REPEAT 3348 3385 PLECTIN 13.  
 FT REPEAT 3386 3423 PLECTIN 14.  
 FT REPEAT 3424 3461 PLECTIN 15.  
 FT REPEAT 3462 3499 PLECTIN 16.  
 FT REPEAT 3500 3537 PLECTIN 17.  
 FT REPEAT 3538 3575 PLECTIN 18.  
 FT REPEAT 3576 3613 PLECTIN 19.

FT REPEAT 3899 3936 PLECTIN 20.  
 FT REPEAT 3937 3974 PLECTIN 21.  
 FT REPEAT 3975 4011 PLECTIN 22.  
 FT REPEAT 4012 4049 PLECTIN 23.  
 FT REPEAT 4050 4087 PLECTIN 24.  
 FT REPEAT 4088 4125 PLECTIN 25.  
 FT REPEAT 4126 4163 PLECTIN 26.  
 FT REPEAT 4164 4201 PLECTIN 27.  
 FT REPEAT 4202 4239 PLECTIN 28.  
 FT REPEAT 4240 4277 PLECTIN 29.  
 FT REPEAT 4278 4315 PLECTIN 30.  
 FT REPEAT 4316 4353 PLECTIN 31.  
 FT REPEAT 4354 4391 PLECTIN 32.  
 FT REPEAT 4392 4429 PLECTIN 33.  
 FT DOMAIN 4253 4303 BINDING TO INTERMEDIATE FILAMENTS.  
 FT DOMAIN 4628 4678 4 X 4 AA TANDEM REPEATS OF G-S-R-X.  
 FT MOD\_RES 4542 4542 PHOSPHORYLATION (BY CDC2) (BY  
 FT SIMILARITY).  
 FT VARSPLIC 1 180 MYAGMLMELDQLRAIYELVFRGVVAVAKDRPRSLPHVP  
 FT MYAGMLMELDQLRAIYELVFRGVVAVAKDRPRSLPHVP  
 FT GVTNLOVWAMTSLSKARGLVRETFAMCHFWYLTNEGIDHL  
 FT ROYLHLPREIYPASIQRYRRPYVAMMPARRSPHYQTMGP  
 FT LGCPKRGPLPAEDPARERORYRKRKEBEGAPETPVSAT  
 FT IVGTLARGPETPAT -> MSQQLRVPEDEGLSKRTSS  
 FT EBNLYAVLRASEGKK (IN ISOFORM 2).  
 FT MYAGMLMELDQLRAIYELVFRGVVAVAKDRPRSLPHVP  
 FT GVTNLOVWAMTSLSKARGLVRETFAMCHFWYLTNEGIDHL  
 FT ROYLHLPREIYPASIQRYRRPYVAMMPARRSPHYQTMGP  
 FT LGCPKRGPLPAEDPARERORYRKRKEBEGAPETPVSAT  
 FT IVGTLARGPETPAT -> MEPSQLPSLVVGHVSLA  
 FT AVHWMRKGRHOADEO (IN ISOFORM 3).  
 FT MYAGMLMELDQLRAIYELVFRGVVAVAKDRPRSLPHVP  
 FT GVTNLOVWAMTSLSKARGLVRETFAMCHFWYLTNEGIDHL  
 FT ROYLHLPREIYPASIQRYRRPYVAMMPARRSPHYQTMGP  
 FT LGCPKRGPLPAEDPARERORYRKRKEBEGAPETPVSAT  
 FT IVGTLARGPETPAT -> DVSNSSGSPSPGDTLPWNLG  
 FT KTORSRSGSGSVGNGSVLDPAERAVIRIA (IN  
 FT ISOFORM 4).  
 FT FT  
 SQ SEQUENCE 4687 AA; 533527 MW; 9966CAF1B929751 CRC64;  
 Alignment Scores:  
 Pred. No.: 8.77e-07 Length: 4687  
 Score: 248.50 Matches: 113  
 Percent Similarity: 44.32% Conservative: 82  
 Best Local Similarity: 25.68% Mismatches: 164  
 Query Match: 9.10% Indels: 81  
 DB: Gaps: 19  
 US-09-502-945-1 (1-1552) x PLEI\_RAT (1-4687)  
 QY 4 CTGATGTCATCGAGAGCTAAACTTACTTATGGAAG-----TGGA 51  
 DB 1675 LeuGlnAlaLeuAspIleuLysLeuGlnAlaGlnAlaGlnArgTrpLeuGlnGln 1694  
 QY 52 ATTGAGATCCCAATGAGTTTGGAGAACGACTTACGTAATCAGAACTTGT 111  
 DB 1695 AlaGlnAlaGlnArgAlaArgAlaGlnValGlnValAlaLeuThrAlaGlnArgSerAla 1714  
 QY 112 GAAGATCTTAAAGAGCACTAAAGCATTAAGATTCTTCTGCTGCTAATCTGTAC 171  
 DB 1715 Glu---ValGlnLeuGlnSerLysArgProSerPhe----- 1725  
 QY 172 CGTGTGGTGGCTTGTGTTGTAATGTCGTCATGGAAGCTGTTTCCCAACCAT 231  
 DB 1726 -----AlaGlnLysThrAlaGlnLeuGlnArgThrLeuGlnGlnLuhis 1740  
 QY 232 ACTAATGTTCAATATG-----CAGACATCGAAGAGCTGTTAAGAAAGAGACTTG 285  
 DB 1741 ValThrValThrGlnLeuArgGlnGlnAlaGlnArgAlaGlnGlnAlaGlnAla 1760  
 QY 286 ATGTCTGCACACTGTTCCGTCGTAAGAGAGCACTGACAGACAAAGAGACAACT 345  
 DB 1761 GluArgAlaArgGlnGlnAlaGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGln 1777

```

OY 346 GCTTATGACAGGTGAACAGTGGTTCGCAATATCTGAGAGACCCATTTGAAAAACC 405
    ||| ||| : : : : ||| ||| : : : : ||| |||
DB 1778 ALaasngluAlaLeuArGleuArGleuGln---AlaGluGluValAlaGlnGlnLysSer 1796
OY 406 AAGCTTTATCCAGTGTCCAGCTTGAGAGACGTGAGCGGAGCCACTT 465
    ||| ||| : : : : ||| ||| ||| ||| ||| ||| |||
DB 1797 LeuAlaGluAlaAspAlaGluLysGlnLysGlnGluAlaGluArGleuAlaArGArg 1816
OY 466 GAAAAAGACCTTCATCTCAGCAGAGAAAGGCGCATTTGAGAAAGCATGATGAAAAAG 525
    ||| ||| : : : : ||| ||| : : : : ||| ||| : : : : ||| |||
DB 1817 GlyLys-----AlaGluGluGlnAlaValArGlnArGleuAlaGluGln 1832
OY 526 GAAATATACGAAAGAGG-----GAGTACATGGGATCAAGATGTTGATCTTGTCT 576
    ||| ||| : : : : ||| ||| : : : : ||| ||| : : : : ||| |||
DB 1833 GluLeuGluLysGlnArGlnLeuThrGlnGluLysThrAlaGlnGlnArGleuAlaAlaGlu 1852
OY 577 CAGAAATATGCCCACTGAGCGCCAGCGTGAA----- 609
    ||| ||| : : : : ||| ||| : : : : ||| ||| : : : : ||| |||
DB 1853 GlnGluLeuLeuArGleuArGleuAlaGluThrGlnGlnGlnLysGlnHisGlnArGlnLeu 1872
OY 610 -----AAGTTTCAAGAAAGAAAGATTTTCACTATTTATCAACTGGAGAA 654
    ||| ||| : : : : ||| ||| : : : : ||| ||| : : : : ||| |||
DB 1873 GluGluGluLeuAlaArGleuGlnHisGlnAlaThrAlaAlaThrGlnLysArGlnGlu 1892
OY 655 ATTCAAGCGCAGCTGCTTCTCGGAAATGATGTCACAAGAGGTGTGTGAGAAATGCCG 714
    ||| ||| : : : : ||| ||| : : : : ||| ||| : : : : ||| |||
DB 1893 LeuGluAlaGluLeuAla-----LysValArGleuAlaGlnGlu 1905
OY 715 TATCAGCTGAAATTAACCAACATGAGAAAGATGAGCGAGAAAGGACACAGAGATTC 774
    ||| ||| : : : : ||| ||| : : : : ||| ||| : : : : ||| |||
DB 1906 ValLeuLeu-----AlaSerLysAlaArGleuAlaGlnGlnLysSerArGSerThr 1921
OY 775 AGACGAAACCACTAACAGCGATCTTGAATTA-----GATCG 813
    ||| ||| : : : : ||| ||| : : : : ||| ||| : : : : ||| |||
DB 1922 SerGlnLysSerLysGlnArGleuGlnAlaGlnAlaGlyArPheArGleuAlaGln 1941
OY 814 GAAATGAGCAATTTGAGATGACTGATGAACCAACACACTTGGACAGAGAGAG 873
    ||| ||| : : : : ||| ||| : : : : ||| ||| : : : : ||| |||
DB 1942 GluAlaAlaArGleuArGleuAlaGlnGlnAlaArGArgHisArGleuAlaGln 1961
OY 874 CAGAGAGCGCCCTGGCGCAGAGAGAGTGCCTGAGA---CTACAGACAGCTCGGGGGA 930
    ||| ||| : : : : ||| ||| : : : : ||| ||| : : : : ||| |||
DB 1962 GluAspAlaAlaArGlnArGlnAlaGlnAlaAspGlyValLeuThrGlnLysLeuAlaAla 1981
OY 931 TCTGACACCACTGCACCTC---ACCAGATCTGAATAGCTCACTCAAGTCAAGAAAA 987
    ||| ||| : : : : ||| ||| : : : : ||| ||| : : : : ||| |||
DB 1982 IleSerGluAlaThrArGleuLysThrGluAlaGlnIleLeu-----LeuLysGlnLys 1999
OY 988 AGGTATACATATGATTAATTTGGAAAGTTA-----CAGACAGAAAT 1029
    ||| ||| : : : : ||| ||| : : : : ||| ||| : : : : ||| |||
DB 2000 GluAlaGluAsnGlnArGleuArGleuAlaGlnAspGluAlaPheGlnArGArg--- 2018
OY 1030 GAAGAATTTGAGAGACAGTGTCCAGCATGGAGATGACATGAGACGATGAAGCAAG 1089
    ||| ||| : : : : ||| ||| : : : : ||| ||| : : : : ||| |||
DB 2019 ---ArGleuGlnGlnAlaAlaGlnHis-----LysAlaAspIleGlnGluArG 2034
OY 1090 CTAAAGCAGCTGTGATGAGACACGACGCGACACGCCAGCAGCTGTGACGCTCTCAGC 1149
    ||| ||| : : : : ||| ||| : : : : ||| ||| : : : : ||| |||
DB 2035 LeuAlaGlnLeuArGlyAlaSerGlnSerGlnLeuGlnArGlnGlnLysGlnLeuVal--- 2053
OY 1150 AAGCAGACACCGTCTCTCGCAGAGAGAGCGCTGTGAGAGAGTGGACCGGCTGGCG 1209
    ||| ||| : : : : ||| ||| : : : : ||| ||| : : : : ||| |||
DB 2054 -----GluAspThrLeuArGlnArGArgGlnValAlaGlnGlnLysIleMetAlaLeuLys 2071

```

```

OS Dictyostellium discoideum (slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX NCBI_TaxId=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=96215148; PubMed=8636147;
RA Hammer J.A. III, Jung G.;
RT "The sequence of the dictyostellium myo J heavy chain gene predicts a
RT novel, dimeric, unconventional myosin with a heavy chain molecular
RT mass of 258 kDa."
RL J. Biol. Chem. 271:7120-7127(1996).
RN [2]
RP SEQUENCE OF 1-1021 FROM N.A.
RX MEDLINE=97039016; PubMed=8884597;
RA Peterson M.D., Urloste A.S., Titus M.A.;
RT "Dictyostellium discoideum myoJ: a member of a broadly defined myosin
RT V class or a class XI unconventional myosin?";
RL J. Muscle Res. Cell Motil. 17:411-424(1996).
RN [3]
RP SEQUENCE OF 182-298 FROM N.A.
RX MEDLINE=95023928; PubMed=7937787;
RA Titus M.A., Kuspa A., Loomis W.F.;
RT "Discovery of myosin genes by physical mapping in Dictyostellium.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9446-9450(1994).
CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN AND HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 TO DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U42409; AAA85186.1; -.
DR EMBL: U35322; AAA79858.1; -.
DR HSSP: P08799; 1MND.
DR DictyDb: D001095; myoJ.
DR InterPro: IPR002710; DIL.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head; 2.
DR Pfam: PF00612; IQ; 6.
DR Pfam: PF01843; DIL; 1.
DR Pfam: PF02736; Myosin_N; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRODOM: PD000355; myosin_head; 2.
DR PRODOM: PD000376; DIL; 1.
DR SMART: SM00015; IQ; 3.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ; 3.
KW Myosin; ATP-binding; Actin-binding; Calmodulin-binding; Repeat;
KW Coiled coil.
FT DOMAIN 1 809 MYOSIN HEAD-LIKE.
FT DOMAIN 824 851 IQ 1.
FT DOMAIN 872 901 IQ 2.
FT DOMAIN 943 972 IQ 3.
FT DOMAIN 973 1812 COILED COIL (POTENTIAL).
FT DOMAIN 1813 2245 TAIL.
FT NP_BIND 174 181 ATP (POTENTIAL).
FT NP_BIND 669 749 ACTIN-BINDING.
FT CONFLICT 191 191 L -> F (IN REF. 2).
FT CONFLICT 284 284 A -> T (IN REF. 2).
FT CONFLICT 291 291 G -> R (IN REF. 2).
FT CONFLICT 332 347 NKGCFEIGVSDDEH -> IEMPELKYRKRS (IN
FT REF. 2).

```



FT CONFLICT 550 550 N -> K (IN REF. 2).  
 FT CONFLICT 865 866 HH -> QQ (IN REF. 2).  
 SO SEQUENCE 2245 AA; 258478 MW; 615E5E1D1A845BE CRC64;  
  
 Alignment Scores:  
 Pred. No.: 1,49e-06 Length: 2245  
 Score: 245.00 Matches: 108  
 Percent Similarity: 37.94% Conservative: 87  
 Best Local Similarity: 21.01% Mismatches: 195  
 Query Match: 8.97% Indels: 124  
 DB: 1 Gaps: 12  
  
 US-09-502-945-1 (1-1552) x MYSJ\_DICDI (1-2245)  
  
 QY 4 CTGATGATCCGAGAACTAAACCTTATGAGAAAGTGAATTCGAGATCC 63  
 DB 1064 LeuaspHisSerLysLeuAsnLysLysLeuGluLysAspLeuSerAspGlnHisAsp 1083  
 QY 64 CAATTGAGTTTGTGAGAACGACTAGCTGAATATCAGAGACTTGTGAGATCTTAA 123  
 DB 1084 SerLiegLysLeuGlnSerGlnPheAsnGlnLthrGlnGlnLthrGlnPheLys 1103  
 QY 124 GAGCACTAAAGCATAAAGAAATTTCTTGCTGCT 159  
 DB 1104 GlnGlnSerGlnGlnLeuSerSerLysLeuSerLysThrGlnGlnLeuAspPheAsn 1123  
 QY 160 -----AATCTGTTAAACCTGCT 177  
 DB 1124 LysGlnGlnPheAspArgLeuSerGlnGlnLArgAspThrAspAsnThrAsnGlnLeu 1143  
 QY 178 GGTGCTGTTGTTGAATGTGCT-----CAGCATGAAGCTGTTCTTCC 222  
 DB 1144 GluLiegGlnLthrLysLysAlaAsnSerThrLeuGlnLysAspLysPheSerLeuSer 1163  
 QY 223 CAAACCCACTAAATGTCATATGCAGACCATCGAA-----AGACTGCTT 267  
 DB 1164 GlyLthrArgAspAsnLeuGlnLArgGlnValLeuGlnLeuArgAspGlnAsnGlnLeuLile 1183  
 QY 268 AAAAGAAAGAGTACTGATGCTGCTGCACTACTTCCGTAAGGAGAGCTTGGA----- 321  
 DB 1184 LysGlnLthrLeuAspSerLeuGlnGlnLThrGlnGlnLThrGlnGlnLThrGlnGlnLThr 1203  
 QY 322 GATACGACGAAAGAGAAAGTGTATGACAGAGTGAAGTGAAGTTTCAATATCT 381  
 DB 1204 GluLysGlnGlnLthrLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1223  
 QY 382 GAGGAGCCAAATTTGAAAAACCAAGGCTTTAATCCAGTGTGACCGAGTGAAGAGAG 441  
 DB 1224 SerGlnLysLeuGlnLysSerGlnGlnLThrGlnGlnLThrGlnGlnLThrGlnGlnLThr 1243  
 QY 442 CTGAGAGAGGCGGCGGAGCGACTTGAAGAAAGACCTGATCGACGAAGAAAGAGGCG 501  
 DB 1244 LeuThrAspHisLysSerLysLeuGlnLThrGlnLthrGlnLthrGlnLthrGlnLthr 1261  
 QY 502 ATTGAGAAAGCATGATGATGAAAGAAATTAACGAAAGAAAGGATACATGGATCAAG 561  
 DB 1262 AsnGlnLys-----LileLysLysLeuLysGlnLysLeuGlnLysGlnLysGlnLys 1279  
 QY 562 ATGTTGATCTTGTCTCAGAAATATGCCCCACTGAGGCCAGGTCGAAAGGTTACAAAG 621  
 DB 1280 -----LysGlnLthrGlnGlnLthrGlnGlnLthrGlnGlnLthrGlnGlnLthr 1291  
 QY 622 GAAAAGATTTACGCTATTATCACTGAGAGAAATTCACACCGACTGCTTCGCGGAA 681  
 DB 1292 SerLysGlnSerValGlnAspGlnLysAsnSerLeuLthrGlnLthrGlnLthrThrValLys 1311  
 QY 682 ATGATGATCAAGAAAGTGTGAGAAATGCGCTATCACTGCAATTAACCAACATGAGAG 741  
 DB 1312 PheGlnSerThrGlnValSerThrAsnValSerHisGlnLysGlnLysLileThrThrLeu 1331  
 QY 742 AAGATGAGGAGCAAGAAAGACACAGAGAGTTCAGAGCAAAACATACAGGAGATCTGAA 801  
 DB 1332 LysSerThrLileGlnGlnLthrLysSerLile---GlyLysLeuGlnAlaGlnLthrLys 1350

QY 802 ATTAAGATCAGAAATAGAGAAATTTGGAATPAGAACTGATGAAGCAACACAC--- 858  
 DB 1351 AsnLysAspAspGlnLileArgLysLileGlnPheGlnLeuAsnAspGlnLysGlnLthr 1370  
 QY 858 ----- 858  
 DB 1371 ThrArgGlnThrLysGlnPheSerAspLeuGlnSerGlnGlnLthrLysSerAspPheGln 1390  
 QY 859 -----TTGGAA 864  
 DB 1391 SerGlnLthrLileHisSerLeuGlnLArgThrAsnGlnLthrLysSerAspPheGln 1410  
 QY 865 CAGGACGACGAGAGCGACCGCTGCGCCACAGAGAGCTC----- 903  
 DB 1411 ArgValGlnGlnSerLeuLysGlnGlnLthrLysGlnLthrLysGlnLthrLysGlnLthr 1430  
 QY 904 -----CTGAGACTAAGAACTGCTGGCGAATCTGAGCACCAG 942  
 DB 1431 AsnArgLeuGlnAsnGlnValLysGlnLthrGlnLthrGlnLthrLysGlnLthrPheGln 1450  
 QY 943 CTGCACCTCACCAGATCTGGAATAGCTCACTCACTCAAGAAAAAGTATACATATGAT 1002  
 DB 1451 PhePheValAlaLysGlnGlnAsnSerAsnGlnLthrGlnGlnLthrLysGlnLthr 1470  
 QY 1003 AAATTGGGAAAGTATACAGAGAAAGAAATGAGAAATGAGGAGGAGGAGTGTCCAGCATGG 1062  
 DB 1471 ValThrThrGlnMetGlnGlnAsnGlnSerArgLileGlnLArgGlnLeu----- 1486  
 QY 1063 AGAGTACATGAGAGCATGAGCAAGCAAGGCTAAGGCGTGAAGGACAGCAGCAGCCACA 1122  
 DB 1487 -----GluGlnLysLysGlnHisLileThrArgLileAspAspGlnLArgAspGlnLeu 1503  
 QY 1123 GCCACGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1182  
 DB 1504 LysLysGlnLthrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1523  
 QY 1183 CTGTCGGAAGAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1242  
 DB 1524 AlaGlnAsnGlnLthrGlnLthrLys----- 1533  
 QY 1243 TGACCTGATGGAACGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1284  
 DB 1534 -----LysGlnLthrLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1540  
  
 RESULT 8  
 MYH3\_CHICK STANDARD; PRT; 1940 AA.  
 ID MYH3\_CHICK  
 AC P02565;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myosin heavy chain, fast skeletal muscle, embryonic.  
 GN MYH3.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87194881; PubMed=3571266;  
 RA Molina M.I., Kropp K.E., Gulick J., Robbins J.;  
 RT "The sequence of an embryonic myosin heavy chain gene and isolation  
 of its corresponding cDNA."  
 RL J. Biol. Chem. 262:6478-6488(1987).  
 RN [2]  
 RP SEQUENCE OF 1502-1940 FROM N.A.  
 RX MEDLINE=83161144; PubMed=6833296;  
 RA Karyinsky C.J., Umada P.K., Sinha A.M., Elzinga M., Tong S.W., Zak R.,  
 RA Jakovick S., Rabinowitz M.;  
 RT "Cloned mRNA sequences for two types of embryonic myosin heavy chains  
 from chick skeletal muscle. I. DNA and derived amino acid sequence of



Db 1574 LysSerAspIleAspArgIleAlaGluLysAspGluIleAspGluLysArg 1593  
 QY 835 GAA-----CTGGATGAAGCAACACTGTGAACGACGACAGCAAGCA 882  
 Db 1594 AsnHisLeuArgValValAspSerMetClnSerThrLeuAspAlaGluIleArg----- 1611  
 QY 883 GCCCTGGCCAGAGAGACTGCTGAGACTTAACAGAACTG-----CTGGCGCA 930  
 Db 1612 -----SerArgAsnGluAlaLeuArgLeuLysLysMetGluGlyAspLeuAsnGlu 1629  
 QY 931 TCTGAGCACCACACTGCACCTCACCAGATCTGAATAGCTCACTCAGTCAAGCAAAAAAG 990  
 Db 1630 IleGluIleGlnLeuSerHisAlaAsnArgGlnAlaAlaGluAlaGlnLysAsnLeuArg 1649  
 QY 991 TATACATATGATTAATTTGGGAAGTACAG-----AGAAGAAAT 1029  
 Db 1650 AsnThrGlnGlyValLeuLeuLysAspThrClnIleHisLeuAspAlaLeuArgSerCln 1669  
 QY 1030 GAAGAATTGGAGAACTGTCTCCAGCATGGAGA-----GTA 1068  
 Db 1670 GluAspLeuLysGluGlnValAlaIleMetValGluArgArgAlaAsnLeuLeuGlnAlaGlu 1689  
 QY 1069 CATAGAGCATGACGACCAAGCTAAGGACGCTGATTAAGCAGCAGCCAGCCAGCCAG 1128  
 Db 1690 IleGluIleLeuArgAlaAlaLeuGlnGlnThrGluArgSerArgLysValAlaGluGln 1709  
 QY 1129 CAGCTG-----GTGACGCTCCTCAGCAGCAAGCAAGCAAGCAAGCTCTCTCTG 1170  
 Db 1710 GluLeuLeuAspAlaSerGluArgValGlnLeuLeuHisThrGlnAsnThrSerLeuIle 1729  
 QY 1171 GAG---AGCGAGAGCCTGCTCGAAGAGGTGGACCGCGCTCGGACCCAGTTA 1218  
 Db 1730 AsnThrLysLysLysLeuGluSerAspLeuSerGlnIleGlnSerGluMet 1746  
 RESULT 9  
 ID PLE1\_HUMAN STANDARD: PRT: 4684 AA.  
 PLE1\_HUMAN  
 AC Q15149; Q16640; Q15148;  
 DT 16-OCT-2001 (Ref. 40, Created)  
 DT 16-OCT-2001 (Ref. 40, Last sequence update)  
 DT 16-OCT-2001 (Ref. 40, Last annotation update)  
 DE Plectin 1 (PLTN) (PCN) (Hemidesmosomal protein 1) (HD1).  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Placenta;  
 RX MEDLINE=96210632; PubMed=8633055;  
 RA Liu C.-G., Maerker C., Castanon M.J., Hauptmann R., Wiche G.;  
 RT "Human plectin: organization of the gene, sequence analysis, and  
 RL chromosome localization (8q24)."  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).  
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND DISEASE.  
 RX MEDLINE=96312447; PubMed=8696233;  
 RA McLean W.H.I., Pulkkinen L., Smith F.J.D., Rugg E.L., Lane E.B.,  
 RA Bullrich F., Burgesson R.E., Amano S., Hudson D.L., Owaribe K.,  
 RA McGrath J.A., McMahon J.R., Eady R.A.J., Leigh I.M., Christiano A.M.,  
 RA Uitto J.;  
 RT "Loss of plectin causes epidermolysis bullosa with muscular dystrophy:  
 RT cDNA cloning and genomic organization."  
 RT Genes Dev. 10:1724-1735(1996).  
 RN [3]  
 RP VARIANT MD-EBS 1003-GLN-ALA-1005 DEL.  
 RX MEDLINE=97049959; PubMed=8894687;  
 RA Pulkkinen L., Smith F.J.D., Shimizu H., Murata S., Yaoita H.,  
 RA Hachisuka H., Nishikawa T., McLean W.H.I., Uitto J.;  
 RT "Homozygous deletion mutations in the plectin gene (PLEC1) in patients  
 RT with epidermolysis bullosa simplex associated with late-onset  
 RT muscular dystrophy."

RL Hum. Mol. Genet. 5:1539-1546(1996).  
 RP [4]  
 RN VARIANT MD-EBS LEU-429 INS.  
 RX MEDLINE=21090821; PubMed=11159198;  
 RA Bauer J.W., Rouan F., Kofler B., Reznicek G.A., Kornacker I.,  
 RA Muss W., Hamelner R., Klausegger A., Huber A., Pohla-Gubo G.,  
 RA Wiche G., Uitto J., Hantner H.;  
 RT "A compound heterozygous one amino-acid insertion/nonsense mutation in  
 RT the plectin gene causes epidermolysis bullosa simplex with plectin  
 RT deficiency."  
 RL Am. J. Pathol. 158:617-625(2001).  
 CC -1- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND  
 CC MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSES OR  
 CC HEMIDESMOSES. COULD ALSO BIND MUSCLE PROTEINS SUCH AS ACTIN TO  
 CC MEMBRANE COMPLEXES IN MUSCLE. MAY BE INVOLVED NOT ONLY IN THE  
 CC CROSSLINKING AND STABILIZATION OF CYTOSKELETAL INTERMEDIATE  
 CC FILAMENTS NETWORK, BUT ALSO IN THE REGULATION OF THEIR DYNAMICS.  
 CC -1- SUBUNIT: HOMODIMER OR HOMOTETRAMER.  
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST LEVELS IN  
 CC MUSCLE, HEART, PLACENTA AND SPINAL CORD.  
 CC -1- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH  
 CC VIMENTIN, DESMIN, GFAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-  
 CC AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.  
 CC -1- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM  
 CC INTERMEDIATE FILAMENTS DURING METOSIS (BY SIMILARITY).  
 CC -1- DISEASE: DEFECTS IN PLECI ARE THE CAUSE OF EPIDERMOLYSIS BULLOSA  
 CC WITH MUSCULAR DYSTROPHY (MD-EBS OR EB-MD), AN AUTOSOMAL RECESSIVE  
 CC DISORDER CHARACTERIZED BY EPIDERMAL BLISTER FORMATION AT THE LEVEL  
 CC OF THE HEMIDESMOSE AND ASSOCIATED LATE-ONSET MUSCULAR DYSTROPHY.  
 CC -1- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 3 PLECTRIN REPEATS.  
 CC -1- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.  
 CC -1- SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: Z54367; CA91196.1; -;  
 DR EMBL: U53204; AAB05427.1; -;  
 DR EMBL: U63610; AAB05428.1; -;  
 DR EMBL: U63609; AAB05428.1; JOINED.  
 DR EMBL: X97053; CAA65765.1; -;  
 DR HSSP: Q01082; 1BKR  
 DR Genew: HGNC:9069; PLEC1.  
 DR MIM: 601282; -;  
 DR MIM: 226670; -;  
 DR InterPro: IPR001589; Actbind\_actin.  
 DR InterPro: IPR001715; Calponin-like.  
 DR InterPro: IPR001101; Plectin\_repeat.  
 DR InterPro: IPR005326; S10\_Plectin.  
 DR InterPro: IPR002017; Spectrin.  
 DR Pfam: PF00307; CH; 2.  
 DR Pfam: PF00681; Plectin; 19.  
 DR Pfam: PF03501; S10\_Plectin; 1.  
 DR SMART: SM00033; CH; 2.  
 DR SMART: SM00250; PLEC; 33.  
 DR SMART: SM00150; SPEC; 5.  
 DR PROSITE: PS00019; ACTININ\_1; FALSE\_NEG.  
 DR PROSITE: PS00020; ACTININ\_2; FALSE\_NEG.  
 DR PROSITE: PS50021; CH; 2.  
 KW Coiled coil. Repeat: Structural protein: Cytoskeleton; Actin-binding;  
 KW Phosphorylation; Alternative splicing; Epidermolysis bullosa;  
 KW Disease mutation.  
 FT DOMAIN 1 1470 GLOBULAR 1.  
 FT DOMAIN 1471 2755 CENTRAL FIBROUS ROD DOMAIN.

FT	DOMAIN	2756	4684	GLOBALAR 2.
FT	DOMAIN	175	400	ACTIN-BINDING.
FT	DOMAIN	179	282	CH 1.
FT	DOMAIN	295	397	CH 2.
FT	REPEAT	645	710	SEPCRIN 1.
FT	REPEAT	740	824	SPCRIN 2.
FT	REPEAT	837	930	SEPCRIN 3.
FT	REPEAT	1315	1415	SEPCRIN 4.
FT	DOMAIN	1469	2756	COILED COIL (POTENTIAL).
FT	REPEAT	2826	2863	PLECTIN 1.
FT	REPEAT	2864	2901	PLECTIN 2.
FT	REPEAT	2902	2939	PLECTIN 3.
FT	REPEAT	2940	2977	PLECTIN 4.
FT	REPEAT	2981	3015	PLECTIN 5.
FT	REPEAT	3116	3153	PLECTIN 6.
FT	REPEAT	3154	3191	PLECTIN 7.
FT	REPEAT	3192	3229	PLECTIN 8.
FT	REPEAT	3230	3267	PLECTIN 9.
FT	REPEAT	3268	3305	PLECTIN 10.
FT	REPEAT	3306	3343	PLECTIN 11.
FT	REPEAT	3485	3522	PLECTIN 12.
FT	REPEAT	3523	3560	PLECTIN 13.
FT	REPEAT	3561	3598	PLECTIN 14.
FT	REPEAT	3599	3636	PLECTIN 15.
FT	REPEAT	3640	3674	PLECTIN 16.
FT	REPEAT	3820	3857	PLECTIN 17.
FT	REPEAT	3858	3895	PLECTIN 18.
FT	REPEAT	3896	3933	PLECTIN 19.
FT	REPEAT	3934	3971	PLECTIN 20.
FT	REPEAT	3975	4008	PLECTIN 21.
FT	REPEAT	4063	4100	PLECTIN 22.
FT	REPEAT	4101	4138	PLECTIN 23.
FT	REPEAT	4139	4176	PLECTIN 24.
FT	REPEAT	4177	4214	PLECTIN 25.
FT	REPEAT	4218	4252	PLECTIN 26.
FT	REPEAT	4265	4305	PLECTIN 27.
FT	REPEAT	4319	4356	PLECTIN 28.
FT	REPEAT	4408	4445	PLECTIN 29.
FT	REPEAT	4446	4483	PLECTIN 30.
FT	REPEAT	4484	4521	PLECTIN 31.
FT	REPEAT	4522	4559	PLECTIN 32.
FT	REPEAT	4560	4597	PLECTIN 33.
FT	DOMAIN	4250	4300	BINDING TO INTERMEDIATE FILAMENTS (BY SIMILARITY).
FT	MOD_RRS	4625	4640	4 X 4 AA TANDEN REPEATS OF G-S-R-X.
FT	MOD_RRS	4539	4539	PHOSPHORYLATION (BY CDC2) (BY SIMILARITY).
FT	VARSPLIC	1	174	MAAGIMLPDQDRLAIEYVLPREGYVAKKDRPRLPHYP
FT	FT	FT	FT	GTYNLQVRAMASLARGVETEFAMCHFEWYLTNEGIALH
FT	FT	FT	FT	ROYLHLPEEIVASLQVRVAMVMPARSPHYOVNGPL
FT	FT	FT	FT	GSPPRKRGPLTEBQRLTKRLKEVSEFETVPAVTRQILA
FT	FT	FT	FT	RGPPEPAPAT -> MSGEDAEVRAVSDVSNSSGSSPPD
FT	FT	FT	FT	TLPMNIGTORRSGGAGNGSVLDPARAVIRIA (IN
FT	FT	FT	FT	ISOFORM 2 AND ISOFORM 3).
FT	FT	FT	FT	MISSING (IN ISOFORM 3).
FT	FT	FT	FT	L -> LL (IN MD-EBS).
FT	FT	FT	FT	/FTID-VAR-011346.
FT	FT	FT	FT	MISSING (IN MD-EBS).
FT	FT	FT	FT	/FTID-VAR-011347.
FT	FT	FT	FT	S -> F (IN REF. 2).
FT	FT	FT	FT	N -> D (IN REF. 2).
FT	FT	FT	FT	H -> N (IN REF. 2).
FT	FT	FT	FT	I -> V (IN REF. 2).
FT	FT	FT	FT	Q -> R (IN REF. 2).
FT	FT	FT	FT	N -> Y (IN REF. 2).
FT	FT	FT	FT	V -> A (IN REF. 2).
FT	FT	FT	FT	V -> L (IN REF. 2).
FT	FT	FT	FT	L -> V (IN REF. 2).
FT	FT	FT	FT	V -> L (IN REF. 2).
FT	FT	FT	FT	I -> M (IN REF. 2).
FT	FT	FT	FT	A -> T (IN REF. 2).
FT	FT	FT	FT	WIC -> RLR (IN REF. 2).
FT	FT	FT	FT	Q -> E (IN REF. 2).

FT	CONFLICT	1789	1789	L -> A (IN REF. 2).
FT	CONFLICT	1910	1910	K -> R (IN REF. 2).
FT	CONFLICT	2154	2154	K -> N (IN REF. 2).
FT	CONFLICT	2160	2160	S -> R (IN REF. 2).
FT	CONFLICT	2215	2215	R -> Q (IN REF. 2).
FT	CONFLICT	2244	2244	S -> A (IN REF. 2).
FT	CONFLICT	3027	3027	K -> E (IN REF. 2).
FT	CONFLICT	3310	3310	E -> A (IN REF. 2).
FT	CONFLICT	3361	3361	F -> L (IN REF. 2).
FT	CONFLICT	3408	3408	F -> L (IN REF. 2).
FT	CONFLICT	3447	3447	S -> A (IN REF. 2).
FT	CONFLICT	3531	3531	G -> A (IN REF. 2).
FT	CONFLICT	3580	3580	R -> S (IN REF. 2).
FT	CONFLICT	3589	3589	K -> Q (IN REF. 2).
FT	CONFLICT	3596	3596	E -> Q (IN REF. 2).
FT	CONFLICT	3616	3616	N -> H (IN REF. 2).
FT	CONFLICT	3686	3686	V -> A (IN REF. 2).
Alignment Scores:				
Pred. No.:	1.53e-06	Length:	4684	
Score:	244.00	Conservative:	117	
Percent Similarity:	41.89%	Conservative:	74	
Best Local Similarity:	25.66%	Mismatches:	187	
Query Match:	8.93%	Indels:	78	
DB:	1	Gaps:	17	
US-09-502-945-1 (1-1552) x PDL_HUMAN (1-4684)				
QY	22	CTAAACTACTATGAGAAAGTGAATTTAGGAATCCCAATGACTTTTGG--	78	
		:		
Db	1456	LeuArgThrHisTyrSerGluLeuThrThrLeuThrSerGlnTyrIleLysPheIleSer	14757	
QY	79	-----AGAACGATTAGCTGAATATCAGAGAACTTGTGA 114		
Db	1476	GluThrLeuArgArgMetGluGluGluGluArgLeuAlaGluGlnArgAlaGluGlu 1495		
QY	115	GATCTTAAGAGCACTAAAGCATAAAGATTCTCTGGCTGTAATACTGTAAACGT 174		
		:		
Db	1496	-----ArgGluArgLeuAlaGluValaGluAlaAlaLeu----- 1506		
QY	175	GTTGGTGGCTTTGTTGTAAGATGGCTCAGACATGAAGCTGTTCTTCCCAACCATACT 234		
		:		
Db	1507	-----GluLysGlnArgGlnLeuAlaGluAlaAla 1517		
QY	235	AATGTTATATGCAGACATCGAAACACTGTTAAAGAA---AGAGATGACTTGATGTC 291		
Db	1518	GlnAlaLysAlaGlnAla---GluArgGlnAlaLysGlnLeuGlnGlnArgIleGln 1536		
QY	292	GCACTACTAGTCCGTAAGGACACTGGCAGACATGACATGACCAAGAGAAAGCATGCTTAT 351		
		:		
Db	1537	GluValValArgArgGluGluAlaAlaValaLysAlaGlnGlnGlnLysArgSerIleGln 1556		
QY	352	GAACAGGTGAACAAGTTTGTGCAAAATATCTGAG-----GAAGCCATTTTGAAAA 402		
		:		
Db	1557	GluGluLeuGlnGlnLeuArgGlnSerSerGluAlaGluIleGlnAlaLysAlaArgGln 1576		
QY	403	ACCAAGCTTTAATCCAGTGTGACCACTTGAGAGAGACCTGAGAGCGACGCGAGCGA 462		
		:		
Db	1577	AlaGluAla---AlaGluArgSerArgLeuArgIleGlnGlnGlnIleArgValValArg 1595		
QY	463	CTTGAAAAAAGAACTTGATCTACGAA----- 489		
		:		
Db	1596	LeuGlnLeuGlnAlaThrGluArgGlnArgGlnArgGlyAlaGluGluGluLeuGlnAlaLeu 1615		
QY	490	GAGAAAGGCGCATGTAAGAAAGCATGAGAAAGGAATTAAGAAAGAAAGGAGATAC 549		
		:		
Db	1616	ArgAlaArgAlaGluGlnAlaGlnAlaGlnLysArgGlnAlaGlnGlnGlnLysArg 1635		
QY	550	ATGGAATCAAGATGTGTGATCTTGTCTCAGAAATATGCCCACACTGGAGCCGAGT--- 606		
		:		
Db	1636	LeuArgArgGlnValaGlnAspGlnSerGlnArgLysArgGlnAlaGluValaGluLeuAla 1655		
QY	607	-----GAAAGGTTACAAAGAAAGATTTTCAGCTATTATTAACAACGT 648		

```

Db 1656 SeratgVallyslagluAlaagluAlaalaargluysgluArgAlaaleuAlaaleu 1675
Oy 649 GAGGAAATTCAGAGCCAGCTGGCTCTCGGGAATGATGTGCACAAAGTGTGTGAGAA 708
Db 1676 GUGUUGUUAAGUUGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGUAGU 1695
Oy 709 ATGCCTATCAGCTGATAT-----AAACCAACATGAGAGAGAGAGAGAGAGAG 759
Db 1696 ArgAlaagluAlaagluAlaaleuAlaaleuAlaaleuAlaaleuAlaaleu 1715
Oy 760 GAGCAGAGA-----GAGTTTCAGAGCAAAACATTAACAGGATCTTGAATTAAGATCAG 813
Db 1716 GlnserLysArgAlaSerPheAlaIuysThr-----AlaGlnleuGluArgSerleuGln 1734
Oy 814 GAAATAGAGAAATTCAGAAATGACACTGCATGATGAAGCAACAACTTGAGACAGAGCAG 873
Db 1735 GUGUUGUUAAGUUAAGUUAAGUUAAGUUAAGUUAAGUUAAGUUAAGUUAAGUUAAG 1754
Oy 874 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 924
Db 1755 AlaGluAlaGluAlaArgAlaArgGluGluAlaGluAlaGluAlaGluAlaGlu 1774
Oy 925 GGCAGATCTGAGACCACTGCACCTCAGACAGATCTGAAATTA-----CTGCTG 966
Db 1775 AlaasngluAlaLeuAlaArgLeuAlaArgLeuAlaArgLeuAlaArgLeuAlaArg 1794
Oy 967 GCTCAGCTCAGTCAAGAAAAAGATATACATATGATTAATTTGGAAAGTTTACAGAGAGA 1026
Db 1795 AlaGluAlaGluAlaGluAlaGluAlaGluAlaGluAlaGluAlaGluAlaGlu 1812
Oy 1027 AATGAAGAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1086
Db 1813 ArgGlyLysAlaGluGluGluGluAlaValaArgGluAlaArgGluAlaArgGlu 1832
Oy 1087 AGGCTAAGCAGAGCTGATTAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1131
Db 1833 LysGlnArgGlnLeu-----AlaGluGlyThrAlaGlnAlaArgLeuAlaAlaGlu 1849
Oy 1132 -----CTGGTGCAGAGCTCTCTCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1185
Db 1850 GlnGluLeuAlaArgLeuAlaArgLeuAlaArgLeuAlaArgLeuAlaArgLeu 1869
Oy 1186 TCGAAGAGAGTGGAGCGGCTGGAGCCAGCTTACCAGATCCAGCA 1233
Db 1870 GlnGluGluLeuAlaAlaArgLeuAlaArgGluAlaAlaAlaAlaAlaThrGln 1885

RESULT 10
MYH9_HUMAN STANDARD: PRT: 1960 AA.
AC P35579: 060805;
DT 01-JUN-1994 (Rel. 29, Created)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,
DE type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).
GN MYH9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M.,
RA Clump M., Smink L.J., Alinsough R., Almeida J.P., Babbage A.K.,
RA Bauguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgman A.M., Buck D., Burgess J.,
RA Burdill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Dhani P.D., Dockree C., Dodsorth S.J., Durbin R.M., Ellington A.G.,
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,

```

```

RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Panlyn G., Heathcote R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Leverisa M.A., Lloyd C., Lloyd D.M.,
RA Martin I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavlit R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willem L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Mioshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Toshizaki Y., Aoki N., Mitsuyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorfman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malat E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozerisky P., Rohlfing T.,
RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budari M.L.,
RA Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
RA Kim U.J., Shizuya H., Simon M.I., Dunamski J.P., Peyrad M., Kedra D.,
RA Seroussi E., Eransson I., Tapia I., Bruder C.E., O'Brien K.P.,
RA Wilkhuun Y., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA Tilihuun Y., Wright H.;
RT "The DNA sequence of human chromosome 22."
RL Nature 402:489-495(1999).
[2]
RP SEQUENCE OF 1-1337 FROM N.A.
RX MEDLINE=92003925; PubMed=1912569;
RA Tootchaker L.E., Gonzalez D.A., Tung N., Lemons R.S., Le Beau M.M.,
RA Arnout M.A., Clayton L.K., Tenen D.G.,
RT "Cellular myosin heavy chain in human leukocytes: Isolation of 5'
RT cDNA clones, characterization of the protein, chromosomal
RT localization, and upregulation during myeloid differentiation."
RL Blood 78:1826-1833(1991).
[3]
RP SEQUENCE OF 1-715 FROM N.A.
RX MEDLINE=91316803; PubMed=1860190;
RA Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,
RA Gdula D., Adelstein R.S., Weir L.;
RT "Human nonmuscle myosin heavy chains are encoded by two genes located
RT on different chromosomes."
RL Circ. Res. 69:530-539(1991).
[4]
RP SEQUENCE OF 714-1960 FROM N.A.
RX MEDLINE=90138958; PubMed=1967836;
RA Saes C.G., Myers J.C., Shows T.B., Leinwand L.A.;
RT "Human nonmuscle myosin heavy chain mRNA: generation of diversity
RT through alternative polyadenylation."
RL Proc. Natl. Acad. Sci. U.S.A. 87:1164-1168(1990).
[5]
RP VARIANT DFN17 HIS-705.
RX MEDLINE=20489856; PubMed=11023810;
RA Lalwani A.K., Goldstein J.A., Kelley M.J., Luxford W., Castelein C.M.,
RA Mhatare A.N.;
RT "Human nonsyndromic hereditary deafness DFN17 is due to a mutation in
RT nonmuscle myosin MYH9."
RL Am. J. Hum. Genet. 67:1121-1128(2000).
[6]
RP VARIANTS MHA/FTNS/SBS K-93; C-702; C-1165; H-1424 AND K-1841.
RX MEDLINE=20428192; PubMed=10973259;
RA Serri M., Cusano M., Gangarossa S., Caridi G., Bordo D., Lo Nigro C.,
RA Ghigeri G.M., Ravazzolo R., Savino M., Del Vecchio M., d'Apollito M.,
RA Iolascon A., Zelanite L.L., Savola A., Baldini C.L., Norris P.,
RA Magrini U., Bellelli S., Heath K.E., Babcock M., Gluckman M.J.,

```

RA Allprandis E., Bizaro N., Desnick R.J., Martignetti J.A.;  
 RT "Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and  
 RT Sebastian syndromes.";  
 RL Nat. Genet. 26:103-105(2000).  
 RN [71]  
 RP VARIANTS MHA ILE-1155 AND LYS-1841.  
 RX MEDLINE-20428193; PubMed-10973260;  
 RA Kelley M.J., Javien W., Ortel T.L., Korczak J.F.;  
 RT "Mutation of MYH9, encoding non-muscle myosin heavy chain A, in  
 RT May-Hegglin anomaly.";  
 RL Nat. Genet. 26:106-108(2000).  
 CC -1- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,  
 CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND  
 CC CAPING.  
 CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY  
 CC CHAIN SUBUNTS (MHC), 2 ALKALI LIGHT CHAIN SUBUNTS (MLC) AND 2  
 CC REGULATORY LIGHT CHAIN SUBUNTS (MLC-2).  
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
 CC -1- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF MAY-HEGGLIN ANOMALY  
 CC (MHA), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED  
 CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUCOCYTE INCLUSIONS.  
 CC (FTNS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED  
 CC WITH ADDITIONAL ALPOPT-LIKE CLINICAL FEATURES OF SENSORINEURAL  
 CC DEAFNESS, CATARACTS AND NEPHRITIS.  
 CC -1- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF SEBASTIAN SYNDROME  
 CC (SBS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED  
 CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUCOCYTE INCLUSIONS.  
 CC -1- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF AN AUTOSOMAL DOMINANT  
 CC FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DEFA17) WHICH IS  
 CC CHARACTERIZED BY PROGRESSIVE HEARING IMPAIRMENT AND  
 CC COCHLEOSACCULAR DEGENERATION.  
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 IO DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: 282215; CAB05105.1; -  
 DR EMBL: M81105; AAA59888.1; -  
 DR EMBL: M69180; AAA61765.1; -  
 DR EMBL: M31013; AAA36349.1; -  
 DR HSSP: P10587; 1BR2.  
 DR GeneW: HGNC:7579; MYH9.  
 DR MIM: 160775; -  
 DR MIM: 153640; -  
 DR MIM: 155100; -  
 DR MIM: 603622; -  
 DR MIM: 605249; -  
 DR InterPro: IPR000048; IO\_region.  
 DR InterPro: IPR004009; Myosin\_N.  
 DR InterPro: IPR002928; Myosin\_tail.  
 DR InterPro: IPR002017; Spectrin.  
 DR InterPro: IPR001609; myosin\_head.  
 DR Pfam: PF00063; myosin\_head; 1.  
 DR Pfam: PF00612; IO; 1.  
 DR Pfam: PF01576; Myosin\_tail; 1.  
 DR Pfam: PF02736; Myosin\_N; 1.  
 DR PRINTS: PR00193; MYOSINHEAVY.  
 DR ProDom: PD000355; myosin\_head; 1.  
 DR SMART: SM00015; IO; 1.  
 DR SMART: SM00242; MYSC; 1.  
 DR PROSITE: PS50096; IO; 1.  
 DR Myosin: ATP-binding; Calmodulin-binding; Actin-binding;  
 KW Coiled coil; Alkylation; Multigene family; Disease mutation;

KW Deafness. 1 778 MYOSIN HEAD-LIKE.  
 FT DOMAIN 779 808 IO.  
 FT DOMAIN 779 808 COILED COIL (POTENTIAL).  
 FT NP\_BIND 837 1926 ATP (POTENTIAL).  
 FT NP\_BIND 174 181 ACTIN-BINDING.  
 FT DOMAIN 654 676 AKTILATION (SH-1) (POTENTIAL).  
 FT MOD\_RES 694 694 AKTILATION (SH-2) (POTENTIAL).  
 FT MOD\_RES 704 704 N -> K (IN MHA).  
 FT VARIANT 93 93 /FTid=VAR.010791.  
 FT VARIANT 702 702 R -> C (IN FTNS).  
 FT VARIANT 702 702 /FTid=VAR.010792.  
 FT VARIANT 705 705 R -> H (IN DENA17).  
 FT VARIANT 705 705 /FTid=VAR.010793.  
 FT VARIANT 1155 1155 T -> I (IN MHA).  
 FT VARIANT 1155 1155 /FTid=VAR.010794.  
 FT VARIANT 1165 1165 R -> C (IN SBS).  
 FT VARIANT 1424 1424 /FTid=VAR.010795.  
 FT VARIANT 1424 1424 D -> H (IN FTNS).  
 FT VARIANT 1841 1841 /FTid=VAR.010796.  
 FT VARIANT 1841 1841 E -> K (IN MHA).  
 FT CONFLICT 53 55 /FTid=VAR.010797.  
 FT CONFLICT 660 660 EAI -> S (IN REF. 3).  
 FT CONFLICT 869 869 T -> S (IN REF. 3).  
 FT CONFLICT 931 931 T -> M (IN REF. 4).  
 FT CONFLICT 1240 1241 C -> Y (IN REF. 4).  
 FT CONFLICT 1350 1350 KG -> GR (IN REF. 4).  
 FT CONFLICT 1764 1764 E -> EE (IN REF. 2).  
 FT CONFLICT 1771 1771 T -> A (IN REF. 2).  
 FT CONFLICT 1771 1771 S -> G (IN REF. 2).  
 Alignment Scores:  
 Pred. No.: 1.95e-06 Length: 1960  
 Score: 243.00 Matches: 106  
 Percent Similarity: 40.87% Conservative: 91  
 Best Local Similarity: 21.99% Mismatches: 161  
 Query Match: 8.89% Indels: 124  
 Db: 1 Gaps: 17  
 US-09-502-945-1 (1-1552) x MYH9\_HUMAN (1-1960)  
 QY 7 GATGCAATCCGAGGAGCAATTAAGTACTTATGAGAAAG----- 45  
 DB 1409 AspLysLeuGluIuLysThrLysThrArgLeuGlnInGluLeuAspPLeuValAsp 1428  
 QY 46 -----TGTGAATTGAGGAATCCCAATTGAAATT----- 75  
 DB 1429 LeuAspHisGlnArgInserAlaCysAsnLeuLysGlnLysPheAspGln 1448  
 QY 76 -----TTGAGCAACGACTTACGTGAATATGAGCACTTGTGAAGATCTTAA 123  
 DB 1449 LeuLeuAlaGluIuLysThrLysLeuSerAlaLysArgLacGluIuArgAspArgAlaGlu 1468  
 QY 124 GAGCACTAAAGCATTAAGAAATTTCTGTGGCTCTAATCTTAACCGGTGGCGCT 183  
 DB 1469 AlaGluAlaArgGluLysGlu-----ThrLysAla 1478  
 QY 184 CTTTGTGTAATGTGCTGCTGAGCATGAGCTGTCTTCCCAACCCATTAATGTTCA 243  
 DB 1479 LeuSerLeuAlaArgAlaLeuGluIuAlaMetGluInLysAlaGlu----- 1494  
 QY 244 ATGCAGACATCGAAGAGCTGGTTAA-----GAAGAGATGACTTGATGCT 291  
 DB 1495 -----LeuGluArgLeuAsnLysGlnPheArgThrLueGluAspLeuMetSer 1511  
 QY 292 GCACTAGTTCCTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 351  
 DB 1512 SerLysAspAspValGlyLysSerValHisGluLeuGluLysSerLysArgAlaLeuGlu 1531  
 QY 352 GAACAGGTGAACAGT-----TTGCAATATCT 381  
 DB 1532 GlnGlnValGluGluMetLysThrGlnLeuGluLeuGluLysPLeuGlnAlaThr 1551

```

OY 382 GAGGAGCC-----AATTTGAAAAAACCAAGCGTTATC----- 417
    |||:|||
Db 1552 GlusaplatylsleuargleuGlulValAsnleuGlnAlaMetLysAlaGlnPheGluArg 1571
OY 418 -----CACTGTGACCAGCTTGAGGAGAGCGTGGAGCGCAGCG 456
    |||:|||
Db 1572 AspleuGlnIyArGAspGluGlnSerGluGlnLysLysGlnLeuValArgGlnVal 1591
OY 457 GAGGAGCTGAAAAAGAACTTCATCTCAGCAAGAAAGAGCCATTGAGAAAGACATG 516
    |||:|||
Db 1592 ArgGluMetGluAlaGluLeuGlnAspGluArgLysGlnArgSerMetAlaValAlaAla 1611
OY 517 ATGAAAAAGCAATACGAAAGAAAGGAGTACATGGATCAAGATCTGATCTGCT 576
    |||:|||
Db 1612 ArgLysLys-----LeuGlu 1616
OY 577 CAGAAATTTGCCCACTGAGCGCCAGTGGAAAAAGTTCAAAAGAAAGATTTCAGCT 636
    |||:|||
Db 1617 MetAspleuLysAspLeuGlnAlaHisIleAspSerAlaAsnLysAsnArgAspGluAla 1636
OY 637 ATTAATCAACTGAGGAAATTCAAAGCAGCTG-----GCTTCTCGGAAATGATGCTC 690
    |||:|||
Db 1637 IleLysGlnLeuArgLysLeuGlnAlaGlnMetLysAspCysMetArgGluLeuAspAsp 1656
OY 691 ACAAGGTGTGTGGAGAAATGCGCTATCAGCTGAATAAACCAAC----- 735
    |||:|||
Db 1657 ThrArgAlaSerArgGluGlnIleLeuAlaGlnAlaLysGlnLysLysLysLys 1676
OY 736 -----ATGAGAGAGATGAGCGCAAGAAAGAGCAGCAGAGAG 771
    |||:|||
Db 1677 SerMetGluAlaGluMetIleGlnLeuGlnGlnIleuAlaAlaGluArgAlaLys 1696
OY 772 TTCAGCAAAAATACTAGAGGATCTGAATTAAGAT----- 810
    |||:|||
Db 1697 ArgGlnAlaGlnGlnGlnLysAsp---GluLeuAlaAspGluIleAlaAsnSerSercly 1715
OY 811 -----CAGAAATAGAGAAATTTGAGAAATGAGAAATGAGTGAAGCAAA 852
    |||:|||
Db 1716 LysGlnLysAlaLeuAlaLeuGlnLysArgArgGlnGlnLysAlaArgIleAlaGlnLeuGlu 1735
OY 853 CAACACTTGGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGA 912
    |||:|||
Db 1736 GluLysLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1755
OY 913 ACAGAACTGCTGGGAGATCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGA 972
    |||:|||
Db 1756 AsnLeuGlnIleAspGlnIleAsnThrAspleuAsnLeuGlnArgSerHis---AlaGln 1774
OY 973 CTCAGTCAGAAAAAGGTATATCATATGATTAATTTGGAAAGTACAGAGAAATGCA 1032
    |||:|||
Db 1775 LysAsnGlnAsnAlaArgGln-----GlnLeuGlnArgGlnAsnLys 1788
OY 1033 GAATTGGAGGAGACAGTGTCTCAGCATGGGAGATACATGAGCAGATGAGCAAAAGCTA 1092
    |||:|||
Db 1789 GluLeuLys-----ValLysLeuGlnGlnLysMetGlnGlnGlnGlnGlnGlnGln 1805
OY 1093 AGCAGCTGATAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGA 1152
    |||:|||
Db 1806 LysAlaSerIleThrAlaLeuGlnAlaLysIleAlaGlnLeuGlnGlnGlnGlnGlnGln 1825
OY 1153 CAGAACAGCTTCTCTGAGAGAGCAGCAGCAGTGTGGAAGAGTGGACCGGCTCGGAGCC 1212
    |||:|||
Db 1826 GluThrLys-----GluArgGlnAlaAlaLysLysGlnValArgArgThrGluLys 1842
OY 1213 CAGCTTA 1218
    |||:|||
Db 1843 LysLeu 1844

```

```

DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Centromeric protein E (CENP-E protein).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis";
RL Nature 359:536-539(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT microtubule motor.";
RL EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and HUBB1.";
RL J. Cell Biol. 143:49-63(1998).
CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBB1 KINASE.
CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CC CONGRESSION. RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z15005; CAAT8727.1; -.
DR PIR; S28261; S28261.
DR HSPG; P17119; 3KAR.
DR Genew; HGNC:1856; CENPE.
DR MIM; 117143; -.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin_1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN; 1.
DR PROSITE; PS00607; KINESIN_MOTOR_DOMAIN; 1.
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere.
FT DOMAIN 1 335 KINESIN_MOTOR.
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NP_BIND 86 93 ATP (BY SIMILARITY).
SQ SEQUENCE 2663 AA; 312087 MW; CERC13880C8C8CB8 CRC64;

```

## Alignment Scores:

```

Pred. No.: 2,12e-06 Length: 2663
Score: 242.00 Matches: 112
Percent Similarity: 43.45% Conservative: 97
Best Local Similarity: 23.28% Mismatches: 162
Query Match: 8.86% Indels: 110
Gaps: 1 22

```



```

US-09-502-945-1 (1-1552) x CENE_HUMAN (1-2663)
OY 1 CTTCTGGAGCATCCGGAAGCTAAACTACTATTATGAGAAAAGTGAATTGAGGAA 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 722 ILEthAspLeuGlnGluLeuAsnLysGluValGlnGlnGlnGlnGlnGlnGln 741
OY 61 -----TCCCAATTGAAGTTTGTGAGAACGACTAGCTGAATATCAGAGA 105
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 742 GluValIleLeuLeuSerGlnLeuLysSerLeuProSerGlnValGlnValGlnGly 761
OY 106 ACTTGTGAGAGATCTTAAAGCACTA-----AAGCATTAAGAATT 147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 762 GluIleGlnAspLysSerGlnGlnLeuHisIleIleThrSerGlnLysAspLysLeuPhe 781
OY 148 CTTCTGGCTCTTAATCTTATACCGTGTGCTGCTTGTGTTGAATGCTGCATCAT 207
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 782 SerGlnValValHisLysGlnSerArgValGlnGlyLeuGlnGlnGlnGlnGln 801
OY 208 GAAGCTGTTCTTCCCAACCCACTACTAAT----- 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 802 LysAspAspLeuAlaThrThrGlnSerAsnTyrLysSerThrAspGlnGlnPheGlnAsn 821
OY 238 -----GTTCAATG----- 246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 822 PheLysThrLeuHisMetAspPheGlnGlnLysTyrLysMetValLeuGlnGlnAsnGln 841
OY 247 -----CAGACCATCGAAGACCTGTAAAGAAAAGACATGACTGTCTGCACCTA 297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 842 ArgMetAsnGlnGlnIleValAlaAsnLeuSerLysGlnValGlnGlnLysPheAspSerLeu 861
OY 298 GTTCCGTAGAAGCAGCCTTGCA-----GATACGACGAGAAAAGAGAAGCACT 345
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 862 GlyAlaLeuLysThrGlnLeuSerTyrLysThrGlnGlnLeuGlnGlnLysThrArgGln 881
OY 346 GCTTATGAACAGCTGAAACACTTTTGCAATATCTGAGAA----- 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 882 ValGlnGlnArgLeuAsnGlnMetGlnGlnLeuLysGlnGlnLeuGlnAsnArgAspSer 901
OY 388 -----GCCAATTTTGAAAAAACCAAGGCTTTAATCCAGTGCAGAGGAGAGAG 441
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 902 ProLeuGlnThrValGlnGlnGlnLysThrLeuIle-----ThrGlnLysLeuGlnGlnThr 920
OY 442 CTGAGAGAGCAGCGAGCGACTTGAAAAAGAA-----CTTCGACTCTCAGCAAG 492
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 921 LeuGln-----GluValLysThrLeuThrGlnGlnLysAspAspLeuLysGlnLeuGlnGln 939
OY 493 AAAAGGCGCATGTGAGAAAGCATGATGAAAAAGAAATACGAAAGAAAGGAGTACATG 552
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 940 SerLeuGlnIleGlnArgAspGlnLeuLysSerAspIleHisAsp----- 954
OY 553 GGATCAAAAGATGTGATCTGTCTCAGAAATAT-----GCCCAACTGAGAGCCCAAG 603
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 955 -----ThrValAsnMetAsnIleAspThrGlnGlnGlnGlnGlnGlnGlnGlnGln 969
OY 604 GTGGAAGAGTTTACAAAGAAAAGATTTCAGTATTATCACTGAGAGAAATTCAAAGC 663
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 970 LeuGlnSerLeu-----LysGlnHisGlnGlnGlnThrIleAsnThrLeuLysSerLysIleSer 988
OY 664 CAGCTGCTTCTCGGGAATGATGTCACAAAGCTGTGTGGAATAATGCC-----TAT 717
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 989 GlnGlnValSerArgAsnLeuHisMetGlnGlnAsnThrGlnLysThrLysAspGlnPhe 1008
OY 718 CAGCTGAATTAACCAACTGAGAGATGAGTACGAGAAAAGAGACAGAGAGTTCAG 777
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1009 GlnGlnLysMetValGlnLysLysGln-----AspLeuGln 1022
OY 778 GCAAAAACATCAACAGGAGCTT-----GAATTAAGATCAGGAATA-----GAG 822
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1023 AlalysAsnThrGlnThrLeuThrAlaAspValLysAspAsnGlnIleIleGlnGlnGln 1042
OY 823 AAATTAGATAGACTGTAAGCAACCAACACTTGGACAGAGACGACGAGAAAGCA 882
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

---

```

Db 1043 ArgLysIlePheSerLeuIleGlnGlnLysAsnGlnLeu-----GlnGlnMetLeuGlnSer 1061
OY 883 GCCCTGGCCAGAGAGAGAGCTCTGAGACTA-----ACAGAAGCTGTGGC 927
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1062 ValIleIleGlnLysGlnGlnGlnLysThrAspLeuLysGlnAsnIleGlnMetThrIle 1081
OY 928 GAATCTGAGCAGCACCACTGACACCTCACCAGATCGAAATFACCTCAA-----CTC 975
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1082 GluAsnGlnGlnGlnLeuArgLeuGlnGlnLysAspGlnLeuLysLysGlnGlnIleVal 1101
OY 976 AGTCAGAAAAAAGGCTTACATATGATTAATTTGGGAAGTTACAGAGAAATAATGAAGA 1035
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1102 AlaGlnGlnLysAsnHisAlaIleLysGlnGlnGlnLeuSerArgThrCysAspArg 1121
OY 1036 TTGAGAGACAGTGTGTCCAGCATGGAGGAGTACATGACAGCATGAAGCAAGAGCTAAG 1095
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1122 LeuAlaGln-----ValGlnGlnLysLeuLysGlnGlnGlnGlnGlnGlnGln 1134
OY 1096 CAGCTGGATTAAGCAGACAGCCAGCCAGCCAGCAGCTGTGACCTCTCAGCAAGCAG 1155
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1135 GlnLeuGlnGlnLysGlnGln-----GlnLeuLeuAsnValGln 1147
OY 1156 AACGAGCTTCTCTGAGAGAGCAGAGCCTGTGGAAGAGTGGACCGCTGCGGAGCCAG 1215
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1148 GlnGln-----MetSerGlnMetGlnLysLysIleAsnGlnIleGlnAsnLeuLysAsnGln 1166
OY 1216 TTA 1218
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1167 Leu 1167

```

---

```

RESULT 12
PEPL_MOUSE
ID PEPL_MOUSE STANDARD: PRT; 1755 AA.
AC Q9R269: Q9JUZ7; Q9CUN1; Q70231;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Periplakin.
GN PPL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA de Vries-Smits A.M.M., Waldmann V., Burgering B.M.T.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/Sv;
RA RYOO Y.W., Li K., Aho S., Cho B.H., Klement J.F., Uitto J.;
RT "Mouse periplakin: genomic cloning and gene targeting.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1643-1755 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai I., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi I., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gliss C., King D.A., Kocliwa H.,
RA Kuenl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Noridone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

```

RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection.",  
 RL Nature 409:685-690(2001).  
 RN (4)  
 RP SEQUENCE OF 1647-1755 FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RX MEDLINE=98190524; PubMed=9521878;  
 RA Aho S., McLean W.H.I., Li K., Uitto J.;  
 RT "cDNA cloning, mRNA expression, and chromosomal mapping of human and  
 mouse periplakin genes.",  
 RL Genomics 48:242-247(1998).  
 CC -1- FUNCTION: COMPONENT OF THE CORNIFIED ENVELOPE OF KERATINOCYTES.  
 CC MAY LINK THE CORNIFIED ENVELOPE TO DESMOSOMES AND INTERMEDIATE  
 CC FILAMENTS (BY SIMILARITY).  
 CC -1- SUBUNIT: MAY FORM A HOMODIMER OR A HETERODIMER WITH EVPL (BY  
 CC SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH DESMOSOMES AND INTERMEDIATE  
 CC FILAMENTS (BY SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 2 PLECTRIN REPEATS.  
 CC -1- SIMILARITY: BELONGS TO THE PLEKIN OR CYTOLINKER FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AF126834; AAD20642.1; -  
 DR EMBL: AF116523; AAF29436.1; -  
 DR EMBL: AF116519; AAF29436.1; JOINED.  
 DR EMBL: AF116520; AAF29436.1; JOINED.  
 DR EMBL: AF116521; AAF29436.1; JOINED.  
 DR EMBL: AF116522; AAF29436.1; JOINED.  
 DR EMBL: AK014700; BAB29510.1; -  
 DR EMBL: AF013715; AAC40068.1; -  
 DR MCD: MGI:1194896; PPL.  
 DR InterPro: IPR001101; Plectin\_repeat.  
 DR InterPro: IPR002017; Spectrin.  
 DR Pfam: PF00681; Plectin; 1.  
 DR SMART: SM00250; PLEC; 2.  
 DR SMART: SM00150; SPEC; 3.  
 KM Keratinization; Repeat; Coiled coil; Cytoskeleton; Structural protein.  
 FT DOMAIN 16 125  
 FT DOMAIN 182 387  
 FT REPEAT 214 315  
 FT REPEAT 321 483  
 FT REPEAT 503 610  
 FT DOMAIN 611 819  
 FT DOMAIN 883 1644  
 FT REPEAT 1650 1684  
 FT REPEAT 1699 1734  
 FT CONFLICT 166 166  
 FT CONFLICT 592 592  
 FT CONFLICT 648 648  
 FT CONFLICT 671 672  
 FT CONFLICT 689 721  
 FT  
 FT CONFLICT 983 983  
 FT CONFLICT 1325 1325  
 FT CONFLICT 1344 1345  
 FT CONFLICT 1755 AA; 204003 MW; 3FEA343086E4CB8F CRC64;  
 SQ SEQUENCE  
 Alignment Scores:  
 Pred. No.: 2.38e-06  
 Score: 241.50  
 Percent Similarity: 46.04%  
 Best Local Similarity: 22.47%  
 Length: 1755  
 Matches: 102  
 Conservative: 107  
 Mismatches: 162

Query Match: 8.84% Indels: 83  
 DB: 1 Gaps: 17  
 US-09-502-945-1 (1-1552) x PEPL\_MOUSE (1-1755)  
 QY 16 GAGAGCTAAACTTACTTATGAGAAAGTGT---GAAATT----- 54  
 DB 1162 GUAAnAlaLysValValGlnGlnLysValArgGluLeuValArgProAspProLys 1181  
 QY 55 GAGGATCCCAATTGAGTTTGGAGAACGACTTATGCTATATCAGAGAACTTGTA 114  
 DB 1182 AlaguserGluValAlaAsnLeuArgLeuLeuValGlnGlnLysPheArg 1201  
 QY 115 GATCTAAAGACACATTAAGCATTAAGATTTCTTGCGCTTAATACTGTAACTG 174  
 DB 1202 GYAlaGlnGlnGlnLeuLeuLysSerTyrGlnSerGluLeuVala---LeuArgAsnArg 1220  
 QY 175 GTTGTGTCTTGTGTTGAAATGCTCAGCATGAGCTGTTTCCCAAC----- 228  
 DB 1221 GYProGlnValGlnValLysGlnValThrLysGlnValIleLysTyrThrAspPro 1240  
 QY 229 CATCTAATGTTCAATATGACACATCCAA-----AGACTGGTT 267  
 DB 1241 GlnThrGlnGlnGlnLeuGlnArgLeuArgGlnGlnIleMetAspLysThrArgLeuIle 1260  
 QY 268 AAAGAAAGAGTACTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 327  
 DB 1261 GlnArgCysAspLeuGlnIleLysGlnLeuLysGlnGlnIleGlnAlaLeuLysAspThr 1280  
 QY 338 CAGCAAGAGAGCAAGCAAGCTTATGACAGAGTGAACAAGTTTTCATATATCTGAGGAA 387  
 DB 1281 LysPro---GlnValGlnThrArgGlnValValGlnGlnIleLeuGlnPheGlnGlnLysp 1299  
 QY 388 GCCAAT-----TTGAAAAAACCAAGCT 411  
 DB 1300 ProGlnThrLysLysGlnValGlnLysSerLeuArgIleGlnLeuSerGlnGlnLysLys 1319  
 QY 412 TTAATCCAGTGTGACAGCTGAGCAAGAGCTGAGAGAGCGAGCGAGCTTGAAGAAA 471  
 DB 1320 GlnValAspLeuGlnGlnGlnArgAlaSerGlnGlnGlnLysIleLysArgLysLysGln 1339  
 QY 472 GAACCTGCATCTCAGCAAGAGAGAAAGGCCATTTGAGAAAGCATG----- 516  
 DB 1340 GlnLeuAlaGlnGlnArgLysGlnArgValValArgGlnGlnValValGlnIleTyrGlnLysp 1359  
 QY 517 -----ATGAAAAAGCAATATACG-----AAA 537  
 DB 1360 GlnProAspLeuArgAlaGlnValThrAlaPheThrAsnSerIleAspAlaGlnLeuArg 1379  
 QY 538 GAAAGGAGTACATGAGATCAAGAGTGTGATCTGCTCAGAAATATGCCCCAGCTGGAG 597  
 DB 1380 GlnIleAspLysLeuIleValGlnLeuArgArgLeuGlnIleAspArgAlaGlnLeuGln 1399  
 QY 598 GCCCAGGTGAGAAAGTTACCAAGGAA-----AAGATTTCAGCTATTATCAACTG 648  
 DB 1400 ArgGlnLeuGlnGlnLeuGlnArgLysGlnAlaArgAlaAlaGlnLeuLeuVal 1419  
 QY 649 GAGGAATTCAGAGCCAGCTGCTGCGGAAATGATGTCACAAAGCTGTGGAGAA 708  
 DB 1420 GlnArgLeuGlnGlnArgLeuAlaLeuGlnGlnGlnGlnIleValLysThr---GlyGln 1438  
 QY 709 ATGCCCTATCAGCTAATTAACCAACATGAGAGAGT---GAGCGAGAAAGAGAGCAG 765  
 DB 1439 LysValIleThrIleThrGlnLysValValLeuGlnAspProGlnGlnThrArgIleLys 1458  
 QY 766 AGAGAGTTCAGAGCAAAATTAACAGGAT-----CTTGAATTAAGATCAGGAA 816  
 DB 1459 AlaLeuLeuArgAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1478  
 QY 817 ATAGAGAAATTTGAGATGAGAACTGATGAGAAAGCAAACTTGAAGAGAGAGCAGAG 876  
 DB 1479 LeuGlnProLeuArgArgLysLeu-----AlaAlaLeuGlnLysAlaGlnIle 1494



```

CC or send an email to llicense@slsb-slb.ch).
CC -----
DR EMBL: U087231; AAB47555.1; -.
DR EMBL: M16557; AAA48970.1; -.
DR PIR: EX0050; PX0051.
DR PIR: A26821; A26821.
DR PIR: S02082; S02082.
DR PIR: S04501; S04501.
DR PIR: S05515; S05515.
DR PDB: 2MYS; 1I-JAN-97.
DR InterPro: IPR000048; IQ-region.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR002928; Myosin_tail.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head.1.
DR Pfam: PF00612; IQ; 2.
DR Pfam: PF01576; Myosin_tail; 1.
DR Pfam: PF02736; Myosin_N; 1.
DR ProDom: PD000355; myosin_head; 1.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Phosphorylation; Acetylation;
KW Calmodulin-binding; Multigene family; 3D-structure.
FT INT_MET 0 0
FT DOMAIN 1 782 MYOSIN HEAD-LIKE.
FT 822 812 IQ.
FT 838 840 HINGE.
FT DOMAIN 841 1938 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP (POTENTIAL).
FT DOMAIN 657 679 ACTIN-BINDING.
FT 759 773 ACTIN-BINDING.
FT MOD_RES 1 1 ACETYLATION.
FT 35 35 METHYLATION (MONO-).
FT MOD_RES 130 130 METHYLATION (TRI-).
FT MOD_RES 551 551 METHYLATION (TRI-).
FT MOD_RES 755 755 METHYLATION (MONO-).
FT MOD_RES 697 697 ALKYLATION (SH-1).
FT MOD_RES 707 707 ALKYLATION (SH-2).
FT CONFLICT 907 907 C -> Q (IN REF. 7 AND 8).
FT CONFLICT 980 980 L -> F (IN REF. 1).
FT CONFLICT 1343 1343 E -> D (IN REF. 5).
FT CONFLICT 1545 1545 S -> A (IN REF. 5).
FT CONFLICT 1796 1797 HV -> QL (IN REF. 5).
FT CONFLICT 1830 1830 S -> A (IN REF. 5).
FT CONFLICT 1863 1863 I -> V (IN REF. 10).
FT CONFLICT 1929 1931 IHG -> FH (IN REF. 10).
SQ SEQUENCE 1938 AA; 223013 MW; EDD01CEA2681E10F CRC64;

Alignment Scores:
Pred. No.: 2.35e-06 Length: 1938
Score: 241.50 Matches: 106
Percent Similarity: 43.75% Conservaive: 97
Best Local Similarity: 22.84% Mismatches: 170
Query Match: 8.84% Indels: 91
DB: 1 Gaps: 19

US-09-502-945-1 (1-1552) x MYSS-CHICK (1-1938)
QY 25 AACTTACTTATGAGAAAGTGT----- 48
D 993 Lysleuthrlrlysglulyslyslaleuclnglualahlsnglntthrlleuasplre 1012
QY 49 GAATTTGAGATCCCAATTTGAAGTTTGAAGAACGACTTGAATATCAGAGAACT 108
D 1013 GlValglugluasplryValasnthrlleuthrllysalatylsthlylsleuclngl 1032
QY 109 TGTGAAGTCTTTAAAGACACACTA---AAGCATTAAGAATTTCTTCGCTGCTAATACT 165
D 1033 Valasplreleugllyserleuclnglulysleuargmetasplreugluarg 1052
QY 166 TGTAACTGTGTGTGTCTTGTGTGAATGTGCTCAGCATGAAGTGTCTTCCCAA 225

```

```

D 1053 Alalysarglylsleuclnglylaspheulysleualh---Hisasperillewetspleu 1071
QY 226 ACCCTACTAATGTTTCATATGACAGCCATCGAAGACGCTTTAAAGAAAGATGACTTG 285
D 1072 Gluasnasplrs-----GlnleuInspsglulyslylslylsasphleglu 1088
QY 286 ATGTCTGACATAGTTTCCGTAAGAGCAGCTGGCAGATPACGACCAAGAACCAACT 345
D 1089 lleser-----GlnlleleInserylilegluaspsgluglnlaleucllymetln 1105
QY 346 GCTTATGAACAGCTGAAACAA-----GTTTTCGAATATTCAGGAACCCAAAT 393
D 1106 Leuglnlylslyllelysgluleuclnlaarglileglulugluleuclnglulrle 1125
QY 394 TTTGAAAAAAC-----AAGCTTTAATCCAGTGTGACCACTTGAAGAGAGCTG 444
D 1126 Alagluargrthrserargalalysalagluulysnlsargalalaspheuserargluleu 1145
QY 445 GAGAGCGAGCGGAGCGACTTGAA----- 468
D 1146 Glugluillesergluargleuclnglualaglyglalathrlalaleglulleglmet 1165
QY 469 -----AAGAAGCTTCATCTCAGCAAGAAAGAGCCATTGAGAAACATGATG 519
D 1166 AsnlyslsargglualagluPhleglnlysmetargrataspheuglulvalathrleu 1185
QY 520 AAAAAAGAAATACCAAGAA-----AGGAGTCAATGGGATCAAAATGTTGATCTTG 573
D 1186 Glnhlsglualathrlalaleuarglylsnlsalaspserthrlalaglu 1205
QY 574 TCTCAGATATTTGCCCACTGAGGCCAGGCGAAGAGTTACAAAGAAAGATTTCA 633
D 1206 Glygluglnlleaspsaspleuclnaryvallysglnlylsleuglulysglulysserclu 1225
QY 634 GCTATTATTAACCTGAGAGAAATTCAGAGCCAGCTGGCTCT-----CGGGAATG 684
D 1226 Leuylsmetglulilleaspsaspleuaslserasmetgluservalserlyslalalsla 1245
QY 685 GATGTCACAAAGCTGTGTGGAAGAAATGCGCTATCAGCTGAATTAACCAACATGACAG 744
D 1246 Asnleuglulysmecllyargthrleucluaspsgluleuserglulilelythrlysglu 1265
QY 745 GATGAGCAGAGAAAG-----GAGCAGAGAGATTGACAGCAAA 783
D 1266 Gluglnasnglnlarymetilleasnaspleuasnthrglnlathrlalargleuclnhrclu 1285
QY 784 ACT-----AACAGGATCTTGAATTAAGATCAGAAATGAGAAATTG----- 828
D 1286 Thrlyglultryserarglinalagluclulysasplalaleulleserclnleuserarg 1305
QY 829 -----AGAAATGAACTGATGAACCAACCAACACTTGGACAGAG----- 870
D 1306 Glylysglnglyphethrglntleugluleucllylsargynhsleuclnglulclulle 1325
QY 871 CAGCAGAGAGCAGCCCTG-----GCCAGAGAGATGGCTGAGACTA 912
D 1326 Lysalalysasnalaleualahlsalaleuclnserlathargnhsaspcysgluleu 1345
QY 913 -----ACAGAACTGCTGGCGGAGTGCAGACCACTGCACCTC 951
D 1346 Arggluglntryglugluluglulnglulnlaalysgllygluleuclnargalaleuserly 1365
QY 952 ACCAGATCTGAATATGCTCAACTCAGTCAAGAAAAAGCTATACATATGATTAATTGGGA 1011
D 1366 Alaasnserylvalalaglle-----TrrargThrlystrygluthrsalalale--- 1382
QY 1012 AAGTTACAGAGAAATGAAGAAATTTGAGAACAGTGTGCCACATGGAGAGACTACAT 1071
D 1383 -----GlnhrgThrGlugluleuclnglu-----Alalys 1392
QY 1072 GAGAGATGAAGCAAGGCTAAGCAGCTGGATTAAGCACAGCCAGCCACAGCCAGAG 1131
D 1131 -----

```

DB 1393 LysLysLeuAlaGlnArgLeuGlnAspAlaGluGlnHisValGlnAlaValAsnAlaLys 1412  
 QY 1132 CTGCTGACGCTCTCGACGAACAGACACCTCTCTCTGAGAGGAGACCTGTGGAA 1191  
 DB 1413 CysAlaSerLeuGlnLysThrLysGlnArgLeuGlnAsnGlnValGlnAspLeuMetVal 1432  
 QY 1192 GAGGTGACCGC 1203  
 DB 1433 AspValGlnArg 1436  
 RESULT 14  
 MYHA\_HUMAN  
 ID MYHA\_HUMAN STANDARD: PRT: 1976 AA.  
 AC P35380;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-FEB-1996 (Rel. 33, last sequence update)  
 DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain,  
 type B) (Nonmuscle myosin heavy chain-B) (NMHC-B).  
 GN MYH10.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96025307; PubMed=7499478;  
 RA Phillips C.L., Yamakawa K., Adelstein R.S.;  
 RT "Cloning of the cDNA encoding human nonmuscle myosin heavy chain-B and  
 analysis of human tissues with isoform-specific antibodies.";  
 RL J. Muscle Res. Cell Motil. 16:379-389(1995).  
 RN [2]  
 RP SEQUENCE OF 63-722 FROM N.A.  
 RX MEDLINE=91316803; PubMed=1860190;  
 RA Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,  
 RA Gidma D., Adelstein R.S., Weir L.;  
 RT "Human nonmuscle myosin heavy chains are encoded by two genes located  
 on different chromosomes.";  
 RL Circ. Res. 69:530-539(1991).  
 CC -1- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,  
 CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND  
 CAPING.  
 CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY  
 CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2  
 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DB EMBL: M69181; AAA99177.1; -  
 DB PIR: B61231; B61231.  
 DB HSSP: P10587; LBR2.  
 DB Genem: HGNC:7568; MYH10.  
 DB MIM: 160776; -  
 DR InterPro: IPR000048; IQ\_region.  
 DR InterPro: IPR004009; Myosin\_N.  
 DR InterPro: IPR002928; Myosin\_tail.  
 DR InterPro: IPR002017; Spectrin.  
 DR InterPro: IPR001609; myosin\_head.  
 DR Pfam: PF00063; myosin\_head; 1.  
 DR Pfam: PF00612; IQ; 1.  
 DR Pfam: PF01576; myosin\_tail; 1.  
 DR Pfam: PF02736; myosin\_N; 1.

DR PRINTS: PR00193; MYOSINHEAVY.  
 DR Prodom: PD000355; myosin\_head; 1.  
 DR SMART: SM00015; IQ; 1.  
 DR SMART: SM00242; MYSC; 1.  
 DR PROSITE: PS50096; IQ; 1.  
 KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;  
 FT DOMAIN 1 785  
 FT DOMAIN 845 1976  
 FT DOMAIN 845 1976  
 FT MOD\_RES 178 185  
 FT MOD\_RES 701 701  
 FT MOD\_RES 711 711  
 SQ SEQUENCE 1976 AA; 228938 MW; B2BB87FF35EA124F CRC64;

## Alignment Scores:

Pred. No.:	2, 65e-06	Length:	1976
Score:	240.50	Matches:	115
Percent Similarity:	46.01%	Conservative:	104
Best Local Similarity:	24.16%	Mismatches:	148
Query Match:	8.80%	Indels:	109
DB:	1	Gaps:	23

US-09-502-945-1 (1-1552) x MYHA\_HUMAN (1-1976)

QY 7 GATGATCCGAGAGCTAAACTTACTTATGAGGAAAGTGTAATTGAGAAATCCCAA 66  
 DB 1038 AspleuGlnGlnArgLeuLysGlnGlnLysGlnGlnLysGlnGlnLysGlnGlnLys 1057  
 QY 67 TTGAAGTT-----TTGAGACGACTTACTGATATTCAGAGAACT 108  
 DB 1058 ArgLysLeuAspLysGlnThrThrAspLeuGlnAlaGlnGlnGlnGlnGlnGln 1077  
 QY 109 TGTGAAGATCTTAAGACCACTTAAGCAATTAATTTCTTGCGTCT----- 159  
 DB 1078 ILAspLysLeuLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1097  
 QY 160 -----AATCTGTAACCGTGTGGTCTTTGTTGAATGCTCAGCAT 207  
 DB 1098 ArgGlyAspAspLysPheLysLysLysLysLysLysLysLysLysLysLysLys 1116  
 QY 208 GAAGCTGTCTCTCCCAACCCACTAATGTCATATGACGACATCGAAGA----- 261  
 DB 1117 GlnAlaGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1136  
 QY 262 CTGCTTAAGAAAGAGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 321  
 DB 1137 AlaGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1156  
 QY 322 GATACGACGAGCAAGAGACGACGATGCTTATGAA-----CAGGTGAACAGATTGCAATA 378  
 DB 1157 AspThrLeuAspPheThrThrAlaAlaGlnGlnGlnGlnGlnGlnGlnGlnGln 1176  
 QY 379 TCTGAG-----GAGCCATTTTGAAGAAACCAAG-----GCTTAATCCAG--- 420  
 DB 1177 AlaGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1196  
 QY 421 -----TGTACCACTTGAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 456  
 DB 1197 MetArgGlnArgHisAlaThrAlaLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 1215  
 QY 457 GAGCGACTTGAAAAGAACTTGATCTCAGCAGAGAAAAGGCGCATGGAAGAAACATG 516  
 DB 1216 LysArgPheLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1232  
 QY 517 ATGAAAAGCAATTAACGAAAGAGAGAGGAGTACATGGGATCAAGATGTTGATCTGTCT 576  
 DB 1233 ---AsnLysGlnLeuAlaLysGln-----ValLysValLeuGln 1244  
 QY 577 CAGAAATATGCGCAACTGAGCGCCAGGTGAGAAAGATTACAAAGAAAGATTACAGCT 636  
 DB 1245 GlnValLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1259

```

QY 637 ATTAATCACTGAGAGAAATTCAGAACCCAGCTGCTCTCGGAA-----ATGGAT 687
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1260 -----GlnValGlnGlnLeuHisAlaLysValSerGlnGlnAspArgLeuAlaGlu 1277
QY 688 CTCACAAAGGTGTGTGGAAATGCGCTATACAGTAATTAACCAATGAGAGAGAT 747
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1278 LeuAlaGlnLysAlaSerLysLeuGlnAsnGlnLeuAspAsnValSerThrLeuGln 1297
QY 748 GAGGCGAAAAAGGAGACACAGAGAGTTCAGAGAAAACCTAACAGGAGATCTTGAAATTA 807
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1298 GlnAlaGlnLysLysGlyIleLysPhe--AlaLysAspAlaAlaSerLeuGlnSerGln 1316
QY 808 GATCAGAGAAATAGAGAAATTC-----828
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1317 LeuGlnAspThrGlnGlnLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1336
QY 829 AGAATA---GAACCTGATGAGAAACCAACACACTTGGACAGAGAGAGAGAGAGAGAGAG 885
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1337 ArgIleArgGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1356
QY 886 CTGGCCAG-----894
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1357 GlnAlaIleArgLysAsnLeuGlnLysGlnValLeuAlaLeuGlnSerGlnLeuAlaAspThr 1376
QY 895 -----GAGAGAGTGCCTGAGACTAACAGAACTGTCGGCGAATCTGAGCACCAG 942
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1377 LysLysLysValAspAspAspLeuGlnThrIleGlnSerLeuGlnGlnAlaLysLys 1396
QY 943 CTGCACCTCACCAGATCTGAAATAGCTCACTCACTCAAGAAAAAGGTATATATATGAT 1002
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1397 LeuLeuLysAspAlaGlnAlaLeuSerGlnArgLeuGlnGlnLysAlaLeuAlaLysAsp 1416
QY 1003 AAATTTGGAAAGTTACAGAGAGA---AATGAGAAATTTGGAGAGAACTGTCTCCAGCAT 1059
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1417 LysLeuGlnLysThrLysAsnArgLeuGlnGlnGlnLeuAspAspLeuThrVal----- 1434
QY 1060 GGGAGAGTACATGAGACATGATGAGCAAGAGCTAAGCGACGTAAACAGCAGCAGCC 1119
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1435 -----AspLeuAsp---HisGlnArgGln 1441
QY 1120 ACAGCCAGCAGCTGCTGACCTCTCAGCAGCAGAACGACCTTCTCTGAGAGAGCAG 1179
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1442 ValAlaSerAsnLeuGlnLysGlnLysGlnLysPheAspGlnLeuLeuAlaGlnLys 1461
QY 1180 AGCCGTGTG-----GAAGAGGTGGACGCGCTCGGAGCCAG 1215
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1462 SerIleSerAlaArgTyrAlaGlnGlnLysArgAspArgAlaGlnAlaGln 1477

RESULT 15
MYPSP_BRUMA STANDARD: PRT: 880 AA.
AC 001202; P90711;
DT 01-FEB-1996 (Rel. 33, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Paramyosin.
OS Brugia malayi.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Brugia.
OX NCBI_TaxID=6279;
RN [1]
RP SEQUENCE FROM N.A.
RA Langy S., Luguid P., Nicolas L.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 180-263 AND 684-880 FROM N.A.
RA MEDLINE=92131068; PubMed=1775173;
RA Li B., Chandrasekar R., Alvarez R.M., Lifits F., Weil G.J.;
RT "Identification of paramyosin as a potential protective antigen
RT against Brugia malayi infection in birds."
RL Mol. Biochem. Parasitol. 49:315-324(1991).
CC -1- FUNCTION: PARAMYOSIN IS A MAJOR STRUCTURAL COMPONENT OF
CC -1- THICK FILAMENTS ISOLATED FROM INVERTEBRATE MUSCLES.

```

```

CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- SIMILARITY: HIGH, TO MYOSIN HEAVY CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U77590; AAC18613.1; -.
DR EMBL; M63097; AAA27859.1; -.
DR EMBL; M63098; AAA27860.1; -.
DR HSSP; P09372; IDKG.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF01576; Myosin_tail; 1.
KW Coiled coil; Muscle protein; Thick filament;
KW Myosin.
FT DOMAIN 1 34 NONHELICAL REGION (POTENTIAL).
FT DOMAIN 35 859 COILED COIL (POTENTIAL).
FT DOMAIN 860 880 NONHELICAL REGION (POTENTIAL).
FT CONFLICT 189 189 V -> I (IN REF. 2).
FT CONFLICT 197 197 D -> A (IN REF. 2).
FT CONFLICT 251 251 L -> F (IN REF. 2).
FT CONFLICT 255 255 E -> Q (IN REF. 2).
FT CONFLICT 696 696 A -> R (IN REF. 2).
FT CONFLICT 730 736 FRKLRR -> IOEAETVA (IN REF. 2).
FT CONFLICT 825 825 L -> LQ (IN REF. 2).
FT CONFLICT 857 880 HOLLRKKLOROKFFESKNSRDN -> SSVVTKKNASAK
FT IYLEDRO (IN REF. 2).
SQ SEQUENCE 880 AA; 101904 MW; F79A27E642D4A85E CRC64;

Alignment Scores:
Pred. No.: 3,35e-06 Length: 880
Score: 239.50 Matches: 117
Percent Similarity: 41.07% Conservative: 113
Best Local Similarity: 20.89% Mismatches: 163
Query Match: 8.77% Indels: 167
Db: 1 Gaps: 21

US-09-502-945-1 (1-1552) x MYPSP_BRUMA (1-880)
QY 4 CTGATGATGATCGGAGAACTAACTATGAGAAAGTGAATTCAGAGATCC 63
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 133 LeuAspTyrAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 152
QY 64 CAATTTGAAGTTTGGAGAGAGACTTACCTGAAATATCAGAGAACTTGAAGATTTAA 123
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 153 ArgGlnArg---LeuGlnHisGlnValIleGlnLeuThrAlaThrIleAspGlnLeuLys 171
QY 124 GAGCACTTAAGCATTAAGAAATTTCTTGCTGCTAACTTTGTAACCTGTGTGGT 183
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 172 LysAsp---LysHis-----LeuAlaGlnLysAlaAlaGlnArgPheGlnAla 186
QY 184 CTTTGTGGAATGTGCTCAGCATGACCTGTCTTCCCAACCACTTAATTTTCAT 243
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 187 GlnThrValGlnLeuSerAsnLysValGlnLysPheAsnArg---HisValAsnLeu 205
QY 244 ATGCAGACCATCGAAGAGAGCTGTTAAAGAGAGATGATGATGCTGACACTAGTTCC 303
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 206 AlaGlnGlnArgGlnArgLeuGlnAlaGlnAsnAsnAspLeuLeuLysGlnIleHisAsp 225
QY 304 GTAAGGAGCAGCTGGAGATACGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 363
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 226 GlnLysValGlnLeuAspAsnLeuGlnHisValLysTyrGlnLeuAlaGlnGlnLeu 245
QY 364 CAAGTTTTCAGAAATATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 423
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 246 GlnAlaIleArgArgLeuGlnLysAlaGlnArgGlnGlnGlnGlnGlnGlnGlnGln 265
QY 424 GACCAAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 483

```





GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 21, 2003, 12:32:38 ; Search time 27.6459 Seconds

(without alignments)

10793.685 Million cell updates/sec

Title: US-09-502-945-1

Perfect score: 2732

Sequence: 1 cttctgcatgcacgcagaaa.....aaatgacctttaagaaga 1552

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p model -DEV=xlp  
-Q/cgr2\_1/USPTO.spool/US09502945/rnatc\_14032003\_101059\_19127/app-query.fasta.1.10979  
-DB=PIR\_73 -QPM=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum2 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFM=pic -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=US09502945 @CGN\_1.1.263 @rnatc\_14032003\_101059\_19127 -NCPU=6 -ICPU=3  
-NO\_XLPRY -NO\_MMAP -LARGEOUTPUT -NRG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-MAIN\_TIMEOUT=30 -HEADS=1 -KAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YCAPOP=10 -YCAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	272.5	10.0	1738	2	T14867
2	264	9.7	1922	2	T00637
3	256.5	9.4	1790	2	S67593
4	255.5	9.4	978	2	A70387
5	255.5	9.4	1959	1	A33977
6	254	9.3	2017	1	A36014
7	254	9.3	2057	2	S61477
8	253.5	9.3	1931	2	A59234
9	253.5	9.3	3225	2	I52300
10	253.5	9.3	3259	1	A56539
11	252	9.2	4574	2	G02520
12	249.5	9.1	2007	1	B43402
13	248.5	9.1	4687	1	A39638
14	246.5	9.0	1992	2	A47297

15	245	9.0	2245	2	T18278	myosin heavy chain
16	244	8.9	764	2	T51302	myosin heavy chain
17	244	8.9	1940	2	A29320	myosin heavy chain
18	244	8.9	4684	2	A59404	plectin [imported]
19	242	8.9	2663	1	S28261	centromere protein
20	241.5	8.8	936	2	S39083	myosin heavy chain
21	241	8.8	1927	2	A59236	embryonic muscle m
22	241	8.8	1940	2	A59287	myosin heavy chain
23	240.5	8.8	1938	1	JX0178	myosin heavy chain
24	240.5	8.8	1976	2	A59252	myosin heavy chain
25	239.5	8.8	746	2	T47237	myosin II heavy ch
26	239.5	8.8	955	2	S24328	myosin heavy chain
27	239.5	8.8	2442	2	T08621	centrosome associa
28	239	8.7	879	2	A48575	paramyosin - nemat
29	239	8.7	1961	1	A61231	myosin heavy chain
30	238.5	8.7	1760	2	T17272	hypothetical prote
31	238	8.7	848	2	A44972	paramyosin - nemat
32	237	8.7	1937	2	I38055	myosin heavy chain
33	236.5	8.7	1934	2	I48153	myosin heavy chain
34	236	8.6	1690	2	T13030	microtubule bindin
35	236	8.6	1935	1	S06006	myosin beta heavy
36	236	8.6	1940	1	A24522	myosin heavy chain
37	235	8.6	1130	2	T34081	hypothetical prote
38	235	8.6	1972	1	A41604	myosin heavy chain
39	234	8.6	741	2	S39082	myosin heavy chain
40	234	8.6	1549	1	A40691	trichohyalin - she
41	234	8.6	1935	2	A59286	myosin heavy chain
42	233.5	8.5	631	2	JC4298	hyaluronan recepto
43	233.5	8.5	924	2	S06117	myosin heavy chain
44	233.5	8.5	1133	2	T22976	hypothetical prote
45	233.5	8.5	1818	1	S73852	hypothetical prote

#### ALIGNMENTS

RESULT 1  
T14867  
Interaplin - slime mold (Dictyostelium discoideum)  
C:Species: Dictyostelium discoideum  
C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 11-May-2000  
C:Accession: T14867  
R:Rivero, F.J.; Kuspa, A.; Brokamp, R.; Matzner, M.; Noegel, A.A.  
J. Cell Biol. 142, 735-750, 1998  
A:Title: Interaplin, an actin-binding protein of the alpha-actinin superfamily in Dicty.  
A:Reference number: Z18248; MUID:98365468; PMID:9700162  
A:Accession: T14867  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1738 <RIV>  
A:Cross-references: EMBL:AF057019; NID:93549260; PID:93549261; PIDN:AMC34582.1  
C:Genetics:  
A:Gene: abpd  
A:Introns: 173/2; 1680/1

#### Alignment Scores:

Pred. No.:	8.96e-10	Length:	1738
Score:	272.50	Matches:	106
Percent Similarity:	42.16%	Conservative:	109
Best Local Similarity:	20.78%	Mismatches:	162
Query Match:	9.97%	Indels:	133
DB:	2	Gaps:	17

US-09-502-945-1 (1-1552) x T14867 (1-1738)

QY	13	TCCGAGAAAGCTAAAGCTTATGAGGAAAGTGTGAATTCGAAATCCCAATGGAAG	72
		:::	
DB	844	ASndltuGASAGncInlAAsphnlsGlnAspGlnleuIeuIuIuSglnleu	863
QY	73	TTTTTGAGGAAGAGCTAGCTGATATCAGAGAACTGTGAAGTCTTAAAGGAACTA	132
DB	864	GlnleuInglngInluyTyrAspGlnleuAsnGlnlthrAsnGlnserIlleGluAsnGlnleu	883

```

OY 133 AAGCATAAAGATTCTTCTGCTGCTAATPACTTGTACCGCTGTGCTGCTTTGTTG 192
DB      :      :      :      :      :      :      :      :      :
DB 884 AaanglInglInLeuInLysInGluAsn-----Leu 894
OY 193 AATATGTCACAGATGACACTGTCTTCCCAAAACCCTACTAATGTTCATATGACAGAC 252
DB      :      :      :      :      :      :      :      :      :
DB 895 AaanglulysglInglInLeuInLysLeuInGlnInLysInGlnInLysInGln 914
OY 253 ATCGAA-----AGACTGTGTAAA 270
DB      :      :      :      :      :      :      :      :      :
DB 915 lIeGlInPheAspGlInglInLupHeserLysGlnInSerLleAsnllIeGluleuValaIn 934
OY 271 GAAAGAGATGACTTGTATGTCGACTAGTTTCCGTAAAGACAGCTTGGACATGACCAAG 330
DB      :      :      :      :      :      :      :      :      :
DB 935 GlulysAsnGlu-----LysLeuInLysInGln 944
OY 331 CAAGAGAAAGCAAGTCTTATGACAGGTAACAAAGTTTGGCAATATCTGAGGAAGC 390
DB      :      :      :      :      :      :      :      :      :
DB 945 Gln-----AspTyrInspGlInLysGlnInGlnAsnInGlnSerInspGlulys 960
OY 391 AAT-----TTGAAAAAACCAAGCCTTATACAGTGTGACAGCTTGAGAGAG 438
DB      :      :      :      :      :      :      :      :      :
DB 961 AspGlulAsnAspLeuInLysGlnInLysInGlnInLysSerLleGlnIn 977
OY 439 GAGCTGAGAGCGAGCGGAGCGACTTGAAAAAGAACTTGCATCTGACGAA---GAGAAA 495
DB      :      :      :      :      :      :      :      :      :
DB 978 GlulEuaSnInglInLeuInLysInGlnInLysInGlnInLysInGlnInLysInGln 997
OY 496 AGGCCCATTTGACAAAGACATGATGAAAAAGAA----- 528
DB      :      :      :      :      :      :      :      :      :
DB 998 GlnSerLleGlInAsnAspLeuInLysGlnInLysGlnInLysGlnInLysGlnIn 1017
OY 529 AAACAGCAAGAAAGAGGACATGAGTGTGATGATGCTGCTGCAATATTTGCC 588
DB      :      :      :      :      :      :      :      :      :
DB 1018 LeuAsnGlnInLysInGlnInLysInGlnInLysInGlnInLysInGlnIn 1034
OY 589 CAATGAGAGCGCCAGCTGAAAAAG----- 612
DB      :      :      :      :      :      :      :      :      :
DB 1035 GlnLeuAsnInglInLeuInLysInGlnInLysInGlnInLysInGlnInLysInGln 1054
OY 613 -----GTTACAAAGAAAGATTTCAAGCTATTAATCACTGAGGAAATTCACAAACCCAG 666
DB      :      :      :      :      :      :      :      :      :
DB 1055 GlnSerLleGlInAsnAspLeuInLysGlnInLysGlnInLysGlnInLysGlnIn 1074
OY 667 CTGGCTTCTCGGGAATGATGTACAAAGTGTGTGAGAAATGCGCTATCAGCTGAAT 726
DB      :      :      :      :      :      :      :      :      :
DB 1075 LeuAsnGlnInLysInGlnInLysInGlnInLysInGlnInLysInGlnInLysInGln 1094
OY 727 AAACCAACATGAGAGATGAGCAGCAAAAGAGACACAGAGTTCAGAGCAAAACT 786
DB      :      :      :      :      :      :      :      :      :
DB 1095 GlnLeu-----lIeGlulysAsnInGlnInLysInGlnInLysInGlnInLysInGln 1113
OY 787 ---AACAGGATCTGAAATTAAGATTCAGAAATGAGAAATTTGGAATAGACAGTGAAT 843
DB      :      :      :      :      :      :      :      :      :
DB 1114 lIeGlulAsnAspLeuInLysGlnInLysInGlnInLysInGlnInLysInGlnInLysInGln 1133
OY 844 GAAAGCAAAACA----- 855
DB      :      :      :      :      :      :      :      :      :
DB 1134 GlulInArgInglInLeuInSerLleAspAsnAspLysLleLeuInGlu 1153
OY 856 -----CACTTGGACAGACAGCAGCAG 876
DB      :      :      :      :      :      :      :      :      :
DB 1154 LysGlnLeuLysGlnInLysInSerLleAspLeuLysLeuAsnAspLulysGlnInGln 1173
OY 877 AAGGAGCCCTGGCCAGAGAGAGTGC-----CTGAGACTATACA----- 915
DB      :      :      :      :      :      :      :      :      :
DB 1174 AspLysGlnLeuInLysInGlnInLysInGlnInLysInGlnInLysInGlnInLysInGln 1193
OY 916 -----GAACTGCTGGCGAATCTGACAGCAACCACTG 945
DB      :      :      :      :      :      :      :      :      :
DB 1194 PheLysAsnAspLysAspSerGlnPheLleGlnInLysInGlnInLysInGlnInLysInGln 1213
OY 946 CACCTCACCAGATCTGAATAGCTCACTCAAGAAAAAGGTATACATAT----- 999

```

```

DB 1214 GlnSerLleGlInLysInLysInGlnInLysInGlnInLysInGlnInLysInGln 1233
OY 1000 -----GATTAATTTGGAAAGTTACAGAGAAAGAAATGAGAGAA 1044
DB 1234 LeuSerGlnLysAspLulysInLysInSerLleGlnInPheGlnInGlnInLysInGln 1253
OY 1045 CAGTGTCTCAGCATGAGAGAGTACATGAGACGATGAAAGGCTAAGCAGTGAAT 1104
DB      :      :      :      :      :      :      :      :      :
DB 1254 GlnLeuSerGlnLysAspLulysInLysInSerLleGlnInLysInGlnInLysInGln 1273
OY 1105 AAGCAGCCAGCCAGCCAGCCAGCAG-----CTGGTGCAGCTC 1143
DB      :      :      :      :      :      :      :      :      :
DB 1274 AspGlulAsnInglInLysValLysGlnInPheSerGlnLysAspLulysInLysInLys 1293
OY 1144 CTCAGACAGCAGACAGCTTCTCCTGAGAGAGCAGAGCTTGGAGAGAGTGCAGCCG 1203
DB      :      :      :      :      :      :      :      :      :
DB 1294 GlnGlnAspLeuAsnInLysInGlnInLysInGlnInLysInGlnInLysInGln 1313
OY 1204 CTGCGGACCCAGTTACCCAGCATGCCCAGAA 1233
DB      :      :      :      :      :      :      :      :      :
DB 1314 LysAspLulysLeuInSerLleGlnIn 1323

```

## RESULT 2

```

T00637
hypothetical protein H_G5541B18.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 05-Nov-1999
C:Accession: T00637
R:Ridms, K.; Sulterer, C.; Becker, M.; Hawkins, M.
submitted to the EMBL Data Library, January 1998
A:Description: The sequence of H. sapiens BAC clone G5541B18.
A:Reference number: 214196
A:Accession: T00637
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1922 <HIN>
A:Cross-references: EMBL:AC004013; NID:g2781380; PIDN:AAB96867.1; PID:g2781381
C:Genetics:
A:Map position: 7q21
A:Note: Intron positions not resolved (incomplete sequence)
A:Note: H_G5541B18.1

```

```

Alignment Scores:
Pred. No.: 3,07e-09 Length: 1922
Score: 264.00 Matches: 121
Percent Similarity: 40.57% Conservative: 107
Best Local Similarity: 21.53% Mismatches: 162
Query Match: 9.66% Indels: 172
DB: 2 Gaps: 22

```

US-09-502-945-1 (1-1552) x T00637 (1-1922)

```

OY 1 CTTCGAGATGATCCGAGAGCTAAACTTACTTATGAGAAAGTGTGAA---ATTGAG 57
DB      :      :      :      :      :      :      :      :      :
DB 51 MetLeuAsnIlleSerSerArgInglInAlaValGlnLysLeuInLysInLysInLys 70
OY 58 GAA-----TCCCAATTAAGTTTGTGAGAAAGCAG 87
DB      :      :      :      :      :      :      :      :      :
DB 71 GlulInSerSerGlnInLysInLysInLysInLysInLysInLysInLysInLysInLys 90
OY 88 TTAGCTGAATATCAGAACTTGTGAAGATCTTAAAGACCACTAAGCATTAAGATTT 147
DB      :      :      :      :      :      :      :      :      :
DB 91 PheArgGlnInLysGlnInLysInLysInLysInLysInLysInLysInLysInLysInLys 110
OY 148 CTTCGCGCTGCTAATACCTTGTAAACCGTGTGCTTTGTTGAAATGTCTCAGCAT 207
DB      :      :      :      :      :      :      :      :      :
DB 111 LeuHISGlnGlnSerArgInLysInLysInLysInLysInLysInLysInLysInLysInLys 128
OY 208 GAAAGCTGCTTCCCAAAACCATACTATATGCTATGACAGCAACCAAGAGAGTGGT 267
DB      :      :      :      :      :      :      :      :      :
DB 129 GlulValIleAspLysLysInLysInLysInLysInLysInLysInLysInLysInLysInLys 146

```



OY	64	CATTGAAGTTTTCGAGACGAGCTTACGTCGATATACAGAACTCTGTAAGACTT---	120
Db	963	LysLeuLysSerLeuAlaAsnAspTyrLysAspMetGlnLacLysAsnSerLeuIle	982
OY	121	-----AAAGAGCACTTAAGACATAAAGAAATTTCTTGTGGCTCTAATGCTGTAAACGT	174
Db	983	LysAlaValGlnGlnSerLysAsnGlnSerSerIleGlnLeuSerAsnLeuGlnAsnLys	1002
OY	175	GTTGGTGGCTCTTGTGTTTGAATATGCTCAGCATGAAGCTGTTCTTCCCAAAACCACT	234
Db	1003	IleAspSer-----MetSerGlnLysGln	1011
OY	235	AATGTCATATGCAG-----ACATCGAAGACTGGTTAAAGAAAGA	276
Db	1012	AsnPheGlnIleGlnArgGlySerIleGlnLysAsnIleGlnLeuLysIleThrIle	1033
OY	277	GATGACTGTGATGTCTGCATAGTTCCTCCGTAAAGACAGACTGGCAGATACGACAAAGA	336
Db	1032	SerAspLeuGlnGlnThrLysGlnGlnIleIleSerLysSerAspSerSerLysAspGln	1055
OY	337	GAGCAGAGTGGTTATGAACAGGTGAACCAAGCTTTGCAATATCTGAGAACCAATTT	396
Db	1052	TyrGlnSerGlnIleSerLeuLeuLysGlnLysLeuGlnThrAlaThrAlaAsnAsp	1077
OY	397	GAAAAAACCAAGCTTTAATCCAGTGTGACAGACTTGAAGAGAGCTGCAGACGAG---	453
Db	1072	GlnAsnValAsnLysIleSerGlnLeuThrLysThrArgGlnGlnLeuGlnLacLeu	1099
OY	454	-----GCGAGGCGACTT	465
Db	1092	AlaAlaTyrLysAsnLeuLysAsnGlnLeuGlnThrLysLeuGlnThrSerGlnLysAla	1111
OY	466	GAAAAAGAACTTGCATCTGCACAGAGAAAGAGGCCCATTTGAGAAACATGATGAAGAAAG	525
Db	1112	LeuLysGlnValLysGlnAsnGlnGlnHisLeuLysGlnGlnLysIleGlnLeuGlnLys	1133
OY	526	GAAATTAAGAAAGAGGAGTACATGTGGATCAAAAGATGTGATCTTGCTCAGAAATAT	585
Db	1132	GlnLacThrGlnThrLysGlnGlnLeuAsnSer-----	1142
OY	586	GCCCAACTGCAGGCCGAGGAGAAAGAGTTACAAAGAAAGATTTCACTTATATCA	645
Db	1143	-----LeuArgAlaAsnLeuGlnSerLeuGlnLysGlnHisGlnLysPheAlaGln	1166
OY	646	CTGAGAGAAATTCAAAGCCAGCTGGCTTCCTCGGAAATGGATGTCAAAAGCTGTGTGA	705
Db	1161	LeuLysLysTyrGlnGlnGlnIleAlaAsnLysGlnArgGlnTyrAsnGlnGlnIleSer	1180
OY	706	GAAATGGCTTACGCTGAATTAACCAACATCGAG-----AAG	744
Db	1181	GlnLeuAsnAspGlnIleThrSerThrGlnGlnGlnLysAsnSerIleLysLysLysAsn	1200
OY	745	GATGAGCGAGAAAGAGCAGACAGAGATTCAGACAAAACTAACAGGATCTTGAAT	804
Db	1201	AspGlnLeuGlnGlnValLysAlaMetLysSerThrSerGlnGlnSerAsnLeu	1220
OY	805	AAAGATCAGAAATAGAGAAATTCAGAAATGAGACTGGATGAAGCAAAACAACACTTGGA	864
Db	1221	LysLysSerGlnIleAspAlaLeuAsnLeuGlnIleLysGlnLeuLysLys-----Lys	1238
OY	865	CAGAGCAGCAGAGGACGACCTG-----GCCAGAGAGAGTGCTG	906
Db	1239	AsnGlnThrAsnGlnLysLeuSerLeuGlnSerIleLysSerValGlnSerGlnThrVal	1255
OY	907	AGACTTAACAGACTGCTGGCGCAATCTGAGCCACCAACTCGACCTCACGATCTGAATA	966
Db	1259	LysIleLysGlnLeu-----GlnAspGlnCysAsnPheLysGlnLysGlnVal	1274
OY	967	GCTCACTCACTCAAGAAAAAGGTATATCATATGATTAATTTGGGAAG-----TTA	1017
Db	1275	SerGlnLeuGlnAspLysLeuLysLysAlaSerGlnAspLysAsnSerLysTyrLeuGlnLeu	1294

OY	1018	CAGGAAGAATGAAAGTAATTGGAGAACAGTGTGCACGANTGGAGAGTCATGAGACG	1077
OY	1295	GlnUlySglUeSerGlUylsIlleYslgUlueuSpalalyStHrHgUleuysIle	1314
OY	1078	ATGAACAAGGCTTAAGCAGCTGATAGCAAGCCAGGCCACAGCCCACAGCTGGTG	1137
Dd	1315	GInLeuGluLysIlleThrAsnLeuSerLySlalysGluLysSerGluSerGluSer	1334
OY	1138	CAGCTTCACAGCACAGACCCTTCTCGAGAGGCGAGAGCTTCGGAAAGAGTG	1197
Dd	1335	AArgLeuLysLySthrSerSerGlu-----GluAArgLysASAlagLUglUGlneu	1351
OY	1198	GACCGGTGGGACCATTTA	1218
Dd	1352	GluLysLeuLysASngIUlle	1358
RESULT 4			
A70387 conserved hypothetical protein ag_1006 - Aquifex aeolicus			
C.Species: Aquifex aeolicus			
C.Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 02-Feb-2001			
C.Accession: A70387			
R.Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;			
Nature 392, 353-358, 1998			
A.Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.			
A.Reference number: A70387; MUID:98196666; PMID:9537320			
A.Status: preliminary; nucleic acid sequence not shown; translation not shown			
A.Molecule type: DNA			
A.Residues: 1-978 <ACF>			
A.Cross-references: GB:A000718; NID:g2983504; PIDN:AMC07092.1; PID:g2983515; GB:AE00			
A.Experimental source: strain VF5			
C.Genetics:			
C.Gene: ag_1006			
C.superfamily: Archaeoglobus fulgidus conserved hypothetical protein AF1032			
Alignment Scores:			
Pred. No.:	1,06e-08	Length:	978
Score:	255.50	Matches:	116
Percent Similarity:	42.16%	Conservative:	107
Best Local Similarity:	21.93%	Mismatches:	157
Query Match:	9.35%	Indels:	149
DB:	2	Gaps:	23
US-09-502-945-1 (1-1552) x A70387 (1-978)			
OY	16	GAGAGCTPAACACTTATGAG-----GAAAAGTGGAAATT	54
Dd	185	IILIIIIIII GlnUleuLysLySGluThrGluLeuLeuLysAspTryThrProHyLysGluVal	204
OY	55	GAGGAATCCCAATTAATTTTGGAGACGCTTAGATATCAGAGAACTTGGAA	114
Dd	205	LeuGluLysThrLeuLysAsnLeuGluGluLeuLysGluLeuLysGluThrGluGlu	224
OY	115	GATCTTAAAGCACTTAACATAAGATTTCCTGCCTGCTTAATCTTGTAAACGT	174
Dd	225	LysLeuAArgGlnGluLeuLysAlaGlu-----	234
OY	175	GTTGGTGGCTTTGTTTAAATGTGCTAGCATGAGCTGTTCTTCCCAAACCATACT	234
Dd	235	-----GluLysAspSerLeuGluAArgGluLeuSerGlnValValThr	248
OY	235	AATGTCATATGACAGACCATGAAAGACTGGTTAAAGAA-----	273
Dd	249	Lys-----LeuLysGluLeuGluasnLeuGluLysGluValGluLysLeuAArgGluLys	266
OY	273	-----	273
Dd	267	L euG l uP heSe rAr gL y sV alAl aPr oHy rAl pr oI leAl aL ysAr giL eG luG luI le	286
OY	273	-----	273

```

Db 287 AsplyslysleuthrGluleuLysValArgLysasnLysleuthrLysGluLeuAlaVal 306
QY 274 ---AGAGATGACTTG-----ATGCTGACAGATTCCGTAGAGAC 312
Db 307 LeuLyspsgluleuSerPheAlaGlnGluLeuAsnArgIleGluAlaGluLysGlu 326
QY 313 AGCTTGGCAGATACGACAAAGAAA-----GCCAGTCTTATGAA 354
Db 327 LysPheLysgluleuLysGluArgGluLysgluleuGlnHisArgLysLysLeuGln 346
QY 335 CAGGTGAACAAGTTTG---CAATATCTGAGACCACTTTGAAAAACCAAGCT 411
Db 347 GluLysLysgluleuLysGluLeuSerLysLeuSerSerLysLysGluLysGlu 366
QY 412 TTAATCCAGTGTGACAGTTGAGAGAGAGAGCGGAGCGGAGCGCTGAA--- 468
Db 367 Arg---GluLysGlnGlnAlaLysGlnGluPheGlnLysPheLysGlnArgValGlnLys 385
QY 469 ---AAGAACTGTGCATCTCAGCAAGAGAAAAGGCCATTGAAAGACATGTAAG 525
Db 386 GlyLysLysLeuValAlaGlnLysGluLys---LeuGlnLys-----LleLys 401
QY 526 GAAATTAACGAAGAAAGAGAGTACGATCAAGATG-----TTGATC 570
Db 402 GluLeuPheSerGlnGluLysGluLysThrSerLeuLysMetLysGluArgLeuValGln 421
QY 571 TTGCTCTGAGATATGCCCCAAGTGT-----GAGGCCAGGTGGAAGGTTACA--- 618
Db 422 LeuGlnArgLysLysGluLysGluLysGluLysGluLysGluLysGluLysGln 441
QY 619 -----AAGAAAAGATP-----TCAGCTATTAATCACTGAGCAAAATTC 660
Db 442 LysTyrLysGluLysLysLysValHisGlnLysValLeuAsnGlnLysGluLysGln 461
QY 661 AGCAGCTGCTTCGCGGAATG-----684
Db 462 ArgGluLysLysGluArgGluLeuHisLysHisAlaHisMetValAlaSerTyrLeuSer 481
QY 685 -----GATGTCACAAAGGTGTGAGAAATGCGCTATGAG-----CTGAAATAA 729
Db 482 ProGlnLysPheLysProValCysGlyGlyLe---TyrArgGlyLysAlaLeuGln 500
QY 730 ACCAATGAGAGAGATGAGGAGCAAGAAAGAGCAGACAGCTTCAAGCAAAATTAAC 789
Db 501 ValAspLysGlnGluLysLeuSerGlnLysHisAlaLysGluLysGluLysGln 520
QY 790 AGGAGCTTGAATTA---GATCAGAA 816
Db 521 ArgGluLysPheLysPheLysLysLysLysLysLysLysLysLysLysLysLys 540
QY 817 ATAGAGAAATTTGAGATAGAACTGATGAAAGCAACACTTGAAACAGAGCAGAC 876
Db 541 MetGluLysLeuArgAsnGlnValGlnGluLeuArgLysGluLe---ProGlnLys 559
QY 877 AAGGACAGCTTGGCAGAGAGAGTGCCTGAGATAACAGAACTGCGGCAATCTGAG 936
Db 560 LysGlnArgLysLysLysLysLysLysLysLysLysLysLysLysLysLys 577
QY 937 CACCAACTGCACCTCACCAGATGTAATAGCTCACTCACTCAAGAAAGATATACA 996
Db 578 HisLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 597
QY 997 TATGATTAATTTGGAGATTAAGATAGAAATGAAATGAGAGCACTGCTGCTCAG 1056
Db 598 GlnAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 617
QY 1057 CATGGAGAGTACATGAGAGATGAAACAA-----AGCTAAGGCGCTGATTAACAC 1110
Db 618 LysSerArgLysValLysGluPheLysGluLeuLysArgValGlnArgLysGlnLys 637
QY 1111 AGCCAGGCCACAGCCAGCAG-----CTGGTCAGAGCTCTCAGC 1149
Db 638 GluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 657

```

```

QY 1150 AAGCAGAACCGCTTCTCTGAGAGCAGAGCCTGTGGAAGAGTGGACCGGCTCGG 1209
Db 658 LysGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 677
QY 1210 ACCCAGTTACCAGCATGCGCAATCT 1236
Db 678 GlyLysLeuSerAlaLeuAsnGlnLys 686

```

## RESULT 5

```

A33977
myosin heavy chain, nonmuscle - chicken
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Gallus gallus (chicken)
C>Date: 31-Dec-1993 #sequence, revision 31-Dec-1993 #text_change 19-Apr-2002
C:Accession: A33977; S06116; A43422
R:Shone, R.V.; Contl, M.A.; Kawamoto, S.; Preston, Y.A.; Brill, D.A.; Adelstein, R.S.
Proc. Natl. Acad. Sci. U.S.A. 86, 7726-7730, 1989
A:Title: Cloning of the cDNA encoding the myosin heavy chain of a vertebrate cellular
A:Reference number: A33977; MUID:90046668; PMID:2813355
A:Accession: A33977
A:Molecule type: mRNA
A:Residues: 1-1959 <SHO>
A:Cross-references: GB:M26510; NID:g212382; PIDN:AAA48974.1; PID:g212383
R:Katsuragawa, Y.; Yanagisawa, M.; Inoue, A.; Masaki, T.
Eur. J. Biochem. 184, 611-616, 1989
A:Title: Two distinct nonmuscle myosin-heavy-chain mRNAs are differentially expressed
s.
A:Reference number: S06116; MUID:90032648; PMID:2806244
A:Accession: S06116
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 716-1008 <KAT>
A:Cross-references: GB:X17589
A:Note: this translation is not annotated in GenBank entry G6MCFMNA, release 114
R:Hodge, T.P.; Cross, R.; Kendrick-Jones, J.
J. Cell Biol. 118, 1085-1095, 1992
A:Title: Role of the COOH-terminal nonhelical tailpiece in the assembly of a vertebra
A:Reference number: A43422; MUID:92381096; PMID:1512291
A:Accession: A43422
A>Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1900-1959 <HOD>
A:Experimental source: brush border
A:Note: sequence extracted from NCBI backbone (NCBI:P.111947)
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleo
F:84-764/Domain: myosin motor domain homology <CMOT>
F:174-181/Region: nucleotide-binding motif A (P-loop)
F:552-565/Region: actin binding #status predicted
F:626-640/Region: actin binding #status predicted
F:837-1936/Domain: coiled coil #status predicted <COI>
F:837-1277/Region: S2
F:1278-1959/Region: light meromyosin
F:1937-1959/Domain: carboxyl-terminal
F:125/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F:180/Binding site: ATP (Lys) #status predicted
F:694,704/Active site: Cys #status predicted

```

## Alignment Scores:

```

Pred. No.: 1,05e-08 Length: 1959
Score: 255.50 Matches: 117
Percent Similarity: 43.93% Conservative: 93
Best Local Similarity: 24.48% Mismatches: 159
Query Match: 9.35% Indels: 109
DB: 1 Gaps: 20

```

US-09-502-945-1 (1-1552) x A33977 (1-1959)

```

QY 1 CTCTGATGATCCGGAAGCTTAATCTTATGAGAAAAGTGAATTTGAGAA 60
Db 1029 TlthrAspLeuGlnGluArgLeuArgGlnGluLysGlnArgGlnGlnGluLys 1048

```

OY	61	TCGCAACTGAGATT-----	-----TTAGGAGACACTTACCATATACG	102
Db	1049	ThraTgaTgTlysLeuGluGlyAspSerSerAspLeuHisAspGlnIleLeuIleGluGln		1066
OY	103	AGAACTGTGAGATCTTAAAGACAACTAAACCATTAAGATTTCTTCTGCTCT---	159	
Db	1069	AlaGlnIleLeuIleGluLeuLysIleGlnLeuSerLysLysGluGluGluLeuGlnIleAlaIa	1088	
OY	160	-----AATACTGTGAACCGGTTGGGGTCTTGTGTTGAAATGTGCTCAG	204	
Db	1089	LeuAlaTgValGluGluGluGlnIleAlaGlnIleLysAsnMetAlaLeuLysLysIleArgGlu	1108	
OY	205	CATGAAGCTGTTCTTTCCCAACCCATACATATGTTTCATATGACAGACCATCAAGAA---	261	
Db	1109	LeuGlnSerGlnIleThrGlnLeuGlnGlnAspLeuGlnSerGlnLysArgLysAsnArgAsn	1128	
OY	262	---CTGCTTAAGAAAGAAAGATGACTTGATGTCTGCACTAGTTTCCGTAAAGAGACGCTTG	318	
Db	1129	LysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGlnIleAlaLeuLysThrIleGlu	1148	
OY	319	GCAGATACG-----	CAGCAAGAAGAGACCACTGCTTATTAACAGAGTG	360
Db	1149	GluAspThrLeuAspSerThrAlaIleGlnGlnGlnLeuArgSerLysArgGluGlnGlu	1166	
OY	361	AAACAGATTTCGCAATATCTGAGGAAGCCAAATTGGAAGAAAAACAGCCTTAATCCAG	420	
Db	1169	ValThrValLeuLysLysThrLeuGluAspGluAlaLysThrHisGlnAlaGlnIleGln	1188	
OY	421	TGTGACGACTGTGAGAGAGAG-----	CTGAGAGCGAGCGGAGGCACTTGAA	468
Db	1189	-----GluMetArgGlnLysHisSerGlnAlaIleGluGluLeuAlaIleGluGlnGlu	1206	
OY	469	AAAGAACTTGCACTCTCAGCAAGAGAAAGGCCATTGAGAAACATGATCAAGAAAGCA	528	
Db	1207	Gln-----	ThrLysArgValLysAlaAsnLeuGlnLys-----AlaLysIleAla	1221
OY	529	ATAAGAAAGAAAGAGGAGTACATGAGTCAAAATGATGTGATCTGTCTGCAGAAATTTGGC	588	
Db	1222	LeuGlnSerGlnLysArgLysGlnLeuSerAsnGlnValLysValLeuGlnGlnIlyGlyGly	1241	
OY	589	CNACTGGAGGCCAGGTGGGAAAGAGTTTACAAAGAAAGATTTCACTATTAATCACTG	648	
Db	1242	AspAlaGlnHisLys-----	ArgLysLysValAlaAspAla-----GlnLeu	1254
OY	649	GAGGAATTCGAACCCAGCGCGCTTCGCGGAAATGATGTCAAAAGGTGTGGAGAA	708	
Db	1255	GlnGluLeuGlnValLysSerThrGlnGluGlyGlu-----	ArgValLysThrGln	1270
OY	709	ATGGCGTATCAGCTGAATTAACCAACATGAG-----	741	
Db	1271	LeuAlaGlnArgValAsnLysLysLeuGlnValGluLeuAspAsnValThrGlyLeuLeuAsn	1296	
OY	741	-----	741	
Db	1291	GlnSerAspSerLysSerIleLysLeuAlaLysAspPheSerAlaLeuGlnSerGlnLeu	1310	
OY	742	AAGATGAGCGAGAAAGAGACACAGAGATGTACAGAAACAAATCAACAGGATCTTGA	801	
Db	1311	GlnAspThrGlnGlnLeuLeuGlnGlnGlnIleThrArgLeuLysLeuSerPheSerThrLys	1330	
OY	802	ATTAAAGATCAGCAATATAGCAAAATTGAGAAATA-----	CNACTGGATGA	846
Db	1331	LeuLysGlnThrThrGlnAspGlyLysAsnAlaLeuLysGlnIleGlnGlnGlnGlnGln	1350	
OY	847	AGCAAAACAACACTTGAAGACAGAG-----	CACCAAGAAGCAAGCCCTGGCCAGA	894
Db	1351	AlaLysArgAsnLeuGlnLysGlnIleSerValLeuGlnGlnGlnAlaValGlnLysArg	1370	
OY	895	GAGGAG-----	TGCCTAGACCTTAACGAAGAACTCTGGGCAATCTGAG	936
Db	1371	LysLysMetLysAspArgLysLeuGlyCysLysGlnIleIleAlaGln-----	GlnValLys	1387
OY	937	CACCAACTGCACCTCACCCAGACTTGGAATATGCTCAACTCAGTCAAGAAAAAGTATACA	996	

Db	1388	LysrLSTAAATGGGAAGCTTACACAGACAGC	---	AATGACGAATTGGAGACACAGCTGTC	1053
QY	997				
Db	1408	TyrspLpLSTAAATGGGAAGCTTACACAGACAGC	---	AATGACGAATTGGAGACACAGCTGTC	1053
QY	1054	CAGCATGGAGACTACATCATGAGACGATGAACAAAGCTAAGCAGCTGATTAAGCAGC	1113		
Db	1428	-----	AspLeuAaSp	-----	HisGln
QY	1114	CAGCCACAGCCCGACGAGTGTGCACGCTCTCAGCAACAGACAGCTTCTCTCGGAG	1173		
Db	1433	ArgInThrValSerAsnLeuGlnLysGlnLysLysPheAspGlnLeuLeuAaGln	1452		
QY	1174	AGGCAGAGCCCTGTG	-----	GAAGAGCTGGACCGCTCGGACCCAG	1215
Db	1453	GlnLysAsnIleSerAlaLysTyrAlaGlnGlnArgAspArgAlaGlnAaGln	1470		

## RESULT 6

myosin heavy chain, nonmuscle - fruit fly (*Drosophila melanogaster*)  
 N:Contains: myosin ATPase (EC 3.6.4.1)  
 C:Species: *Drosophila melanogaster*  
 C:Date: 31-Dec-1993 #sequence, Revision 31-Dec-1993 #text, Change 19-Apr-2002  
 C:Accession: A36014; B36014  
 R:Kerchum, A.S.; Stewart, C.T.; Stewart, M.; Kiehart, D.P.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 6316-6320, 1990  
 A:Title: Complete sequence of the *Drosophila* nonmuscle myosin heavy-chain transcript.  
 A:Reference number: A36014; MUID:90349606; PMID:2117279  
 A:Accession: A36014  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1\*2017 <KET>  
 A:Cross-references: GB:M35012  
 C:Genetics:

A:Gene: FlyBase:zfp  
A:Cross-references: FlyBase:FBgn0005634  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylat  
F:1-2017/Product: myosin heavy chain, form I #status predicted <MA1>  
F:46-2017/Product: myosin heavy chain, form II #status predicted <MA2>  
F:135-815/Domain: myosin motor domain homology <MOT>  
F:125-232/Region: nucleotide-binding motif A (P-loop)  
F:598-631/Region: actin binding #status predicted  
F:705-727/Region: actin binding #status predicted  
F:888-2017/Domain: coiled coil #status predicted <COI>  
F:888-1328/Region: S2  
F:1329-2017/Region: light meromyosin  
F:176/Modified site: N6/N6-crimeThyllysine (lys) #status predicted  
F:231/Binding site: ATP (lys) #status predicted  
F:745,755/Active site: Cys #status predicted

**Alignment Scores:**

Pred. No.:	1.31e-08	Length:	2017
Score:	254.00	Matches:	110
Percent Similarity:	43.158	Conservative:	101
Best Local Similarity:	22.49%	Mismatches:	174
Query Match:	9.30%	Indels:	104
DB:	1	Gaps:	18

US-09-502-945-1 (1-1552) x A36014 (1-2017)

Oy	4	CTGATGTCATCCGAGAGCTAAACCTACTATAGAGAAATGTTGAAATTTAGGAATCC	63
		:::	
Db	1143	IIespjngIuSeratIhrIysalItnIalngIylsIagIlnrgG IuIengIuc	1162
Oy	64	CAATTGAAGTTTGGAGAACGACTTA-----GCTGAATATTCAG	102
Db	1163	GIIneuIaIaIuIeIingIuIuIuIuIaIaIuIyIaIaIaIaIaIaIaIuI	1182
Oy	103	AGAACTGTGAAGACTTTAAAGGACACTAAG-----CATAAGAATTTCTCTGGCT	156
		:::	

```

Db 1183 LysValAlaGArgAspLeuSerGluGluLeuAlaLeuLysAsnGluLeuLeuAspSer 1202
QY 157 GCTAATACCTGTAACCGGTGGTGTCTGTTGAAATGCGTCAG----- 204
Db 1203 LeuAspThrThrAlaAlaGlnGlnGlnLeuArgSerLysArgGluGlnGluLeuAlaThr 1222
QY 205 -----CATGAAGCTGTTCTT----- 219
Db 1223 LeuLysLysSerLeuGluGluGluThrValAsnHisGlnGluLysAlaLeuAlaAspMetArg 1242
QY 220 -----TCCAAACCCCACTAATGTTGATTCAGACCATCGAAAGCTGGTTAA 270
Db 1243 HisLysHisSerGlnGluLeuAsnSerLLeasnAspGln---LeuGluAsnLeuArgLys 1261
QY 271 -----GAAAGAGATGACTGTGATG 288
Db 1262 AlaLysThrValLeuGluLysAlaLysGlyThrLeuGluAlaGlnAsnAlaAspLeuAla 1281
QY 289 TCTGCACCTAGTTTCCGTAAGAGCAGCTTGGCAGATACGCAAGCAAGAGAACTGCT 348
Db 1282 ThrGluLeuArgSerValAsnSerSerArgGlnGlnAsnAspArgArgLysGlnAla 1301
QY 349 TATGACACGCTGAAACAAGTTTGCATAATCTGAGAGCAACCAATTTGAAAAACCAAG 408
Db 1302 GluSerGlnThrAlaGlu---LeuGlnVal---LysLeuAlaGlnLysGluArgLys 1319
QY 409 GCTTAATCCAG-----TGTACCAGTTGAGGAAGCAGCTGAGAGCGCAGCGAGCA 462
Db 1320 SerGluLeuGlnGluLysCysThrLysLeuGlnGlnGlnAlaGlnAsnThrAsnGln 1339
QY 463 CTGGAAGAA-----GAACTTGATCTCCACCAAGCAAA 495
Db 1340 LeuGlnGluAlaGlnLeuLysAlaSerAlaAlaValLysSerAlaSerAsnMetGluSer 1359
QY 496 AGGCCCATGGAAGACATGATGAAAAAGAAATTAACGAAAGAAAGGAGCATGATGCGA 555
Db 1360 GlnLeuThrGlnAlaGlnGlnLeuGlnGlnGlnLysThrArgGlnLys----- 1375
QY 556 TCAGAAAGTGTGATCTGTCAGAAATATGCCCACTGAGCGCCAGCGTGAAGAAAGTT 615
Db 1376 -----LeuGlyLeuSerSerLysLeuArgGlnLysGlnLysGlnAlaLeu 1392
QY 616 ACAAGGAAAAAGATTTCAGCTATTATCACTGAGCAAAATTCAAAGCCAGCTGCTCT 675
Db 1393 GlnGlu-----GlnLeuGlnGlnAspAspGlnAlaLysArgAsn 1405
QY 676 CGGGAATGCAATGCACAAAGCTGTGAGAAATGCGCTATCAGCTGAATAAACCAAC 735
Db 1406 TyrGlnLysLysLeuAlaGlnValThrThrGlnMetGln---GlnLysLysLysAla 1424
QY 736 ATGAGAGAGATGAGCGCAAGAAAGAGACAGACAGATTCAAGCAAGCAAGAGAT 795
Db 1425 GlnGluAspAlaAspLeuAlaLysGlnGlnGlnGlnLysLysArgLeuAsnLysAsp 1444
QY 796 CTGGAATTAAGATCAAGAAATAGAGAAATAGAAATAGAA-----CTGATGAA 846
Db 1445 IlleGlnAlaLeuGlnLysArgGlnValLysLeuLeuAlaGlnAsnAspArgLeuAspLys 1464
QY 847 AGCAAAACAACCTTGAACAGCAGCAGCAGCAGCAGCAGC----- 885
Db 1465 SerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1484
QY 886 -----CTGGCCAGAGAG 897
Db 1485 ThrLysValLeuGlnLysLysLysLysLysAsnAspLysLysLysLysLysLysLys 1504
QY 898 GAGTGCCTG-----AGACTAACAGAACTGCTGGCGAATCTGACCAACTGCACCTC 951
Db 1505 LysAlaThrSerGlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 1524
QY 952 ACCGAGATCTGAATAGCTCAACTCACTCAAGAAAAAGGTTACATATGATTAATGGGA 1011
Db 1525 LysGlnThrLysValLeuSerValSerArgLysLysLysLysLysLysLysLysLysLys 1544

```

```

QY 1012 AAGTTACAGAGAAATGAGAAATTTGAGAGACAGTGTCTCCAGCATGGAGATACAT 1071
Db 1545 AspLeuGlnAsnLysArgLysThrLeuGlnAsnGlnLysLysLysLysLysLysLys 1564
QY 1072 GAGACATGAGCAAGAGCTTAAGCAGCTGATTAAGCAGCAGCAGCAGCAGCAGCAG 1131
Db 1565 GlyThrAlaAspLysAsnValHisGlnLeuGlnLysLysLysLysLysLysLysLys 1584
QY 1132 CTGTGACGCTCTTCAGACAGACAGCAGCTTCTCTGGAGAGACAGCAGCTTCGGA 1191
Db 1585 LeuAlaGlu---LeuLysAlaGlnAsnGlnLysLysLysLysLysLysLysLysLys 1603
QY 1192 GAGGTGACGCGCTGGAGACCCAGTTA 1218
Db 1604 AspAla---LysLeuArgLeuGlnVal 1611

RESULT 7
S61477
myosin II heavy chain, non-muscle - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 02-Feb-2001
C:Accession: S61477; S65349
R:Manuscript, S.G.; Al-Shirawi, D.Y.; Ketchum, A.S.; Newbern, E.C.; Kiehart, D.P.
J. Mol. Biol. 255, 98-109, 1996
A:Title: Molecular organization and alternative splicing in zipper, the gene that enc
A:Reference number: S61477; MUID:96144835; PMID:8568878
A:Accession: S61477
A:Molecule type: DNA
A:Residues: 1-2057 <MAN>
R:Manuscript, S.G.; Al-Shirawi, D.Y.; Ketchum, A.S.; Newbern, E.C.; Kiehart, D.P.
submitted to the EMBL Data Library, September 1995
A:Reference number: S65349
A:Molecule type: DNA
A:Residues: 1-1908, 'N/L', 1911-2057 <MAN>
A:Cross-references: EMBL:U35816; NID:g1141789; PIDN:AA09049.1; PID:g1572481
A:Gene: zip
A:Cross-references: FlyBase:FBgn0005634
A:Interons: 10/3; 90/3; 162/3; 215/1; 255/3; 263/3; 303/3; 334/3; 1248/1; 1349/3; 1526
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: alternative splicing; ATP; nucleotide binding; P-loop
F:135-555/Domain: myosin motor domain homology <MOT>
F:225-232/Region: nucleotide-binding motif A (P-loop)

Alignment Scores:
Pred. No.: 1,31e-08 Length: 2057
Score: 254.00 Matches: 110
Percent Similarity: 43.15% Conservative: 101
Best Local Similarity: 22.49% Mismatches: 174
Query Match: 9.30% Gaps: 104
DB: 2

US-09-502-945-1 (1-1552) x S61477 (1-2057)
QY 4 CTGATGATCGTCCGAGAACTTAATCTTATGAGAAAGTGTGAATTTGAGAAATCC 63
Db 1183 IleAspGlnGlnSerAlaThrLysAlaThrAlaGlnLysAlaGlnArgGlnGlnLys 1202
QY 64 CAATTGAAGTTTGGAGAACGACTTA-----GCTGAATATCAG 102
Db 1203 GlnLeuAlaGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1222
QY 103 AGAAGCTGTGAAGATCTTAAAGAGCAACTAAG-----CATAAAGATTTCTTCTGCT 156
Db 1223 LysValAlaGArgAspLeuSerGlnGlnLysLysLysLysLysLysLysLysLysLys 1242
QY 157 GCTAATACCTGTAACCGCTGTGGCTGCTTTGTTGAAATGCTGCG----- 204
Db 1243 LeuAspThrThrAlaAlaGlnGlnGlnLysLysLysLysLysLysLysLysLysLys 1262

```



```

OY 205 -----CATGACCTGCTCT----- 219
Db 1263 LeuLysSerLeuGluGluThrValAsnHisGluLysAlaAspMetArg 1282
OY 220 -----TCCGAACCCATGATCATATGATGACACATCGAAGACGTTTAA 270
Db 1283 HisLysHisSerGlnLysAsnSerIleAsnAspGln---LeuGluAsnLeuArgLys 1301
OY 271 -----GAAAGATGACTTGATG 288
Db 1302 AlaLysThrValLeuGluLysAlaLysGlyThrLeuGluLysAlaAspLeuAla 1321
OY 289 TCAGCACTACTTCCGTAAGAGCAGCTTGCGACATACGACGACGAGAGAGCAAGTCT 348
Db 1322 ThrGluLeuArgSerValAsnSerSerArgGlnGluAsnAspArgArgLysGlnAla 1341
OY 349 TATGACAGCTGGAACAGTTTGCAGATATCGAGACCCCAATTTGAAAAACCAAG 408
Db 1342 GluSerGlnIleAlaGlu---LeuGlnVal---LysLeuAlaGluIleGluArgAlaArg 1359
OY 409 GCTTTAATCCAG-----TGTGACCACTTGAGAGAGAGCTGAGAGCGAGCGAGCGA 462
Db 1360 SerGluLeuGlnGluLysCysThrLysLeuGlnGlnGluLysAlaAsnIleThrAsnGln 1379
OY 463 CTGAAAAA-----GAACTGCATCTCAGCAAGAGAA 495
Db 1380 LeuGlnGluLysAlaGluLeuLysAlaSerAlaValLysSerIleAsnMetGluSer 1399
OY 496 AGGCACTGAGAAAGCATGATGATAAGAAATAGAAAGAAAGAGAGATGATGGA 555
Db 1400 GlnLeuThrGluLysAlaGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1415
OY 556 TCAAGATGTTGATCTTGTCTCAGAAATATTCGCCAATGAGGCCAGCGGAGAAAGTT 615
Db 1416 -----LeuGlyLeuSerLysLeuArgGlnIleGluSerGluLysGluAlaLeu 1432
OY 616 ACAAGAGAAAGATTTCAGTATTAATCACTGAGAGAAATTCAAACCCAGCTGCTCT 675
Db 1433 GlnGlu-----GlnLeuGlnLysAspArgLysAlaLysArgAsn 1445
OY 676 CGGAAATGATGATGCAAAAGTGTGTGAGAAATGCGCATGATGATTAACCAAC 735
Db 1446 TyrGlnArgLysLeuAlaGluValThrGlnMetGln---GluIleLysLysLysAla 1464
OY 736 ATGAGAGAGATGAGCAGAGAAAGAGACAGAGATTCAGACCAAACTAACAGGAT 795
Db 1465 GlnGluAspIleAspLeuAlaLysGluLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 1484
OY 796 CTGGAATTAATGATCAGCAAAATAGAGAAATTCAGATAGAA-----CTGGATGAA 846
Db 1485 IleGluAlaLeuGlnLysGlnValLysGlnLeuIleAlaGlnAsnAspArgLeuAspLys 1504
OY 847 AGCAACACACTTGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 885
Db 1505 SerLysLysLysIleGlnSerGlnLeuGlnAspAlaThrIleGlnLeuAlaGlnArg 1524
OY 886 -----CTGGCCAGAGAG 897
Db 1525 ThrLysValLeuGlnLysLysGlnLysAsnPheAspLysIleLeuAlaGlnGlu 1544
OY 898 GAGTGCCTG-----AGACTAACAGAACTGCTGGGCAATCTGAGACCAACAGCTCC 951
Db 1545 LysAlaIleSerGlnIleAlaGlnGlnArgAspThrAlaGlnArgLysAlaArgGln 1564
OY 952 ACCAGATCGAATGACTCAACTCAGCAAGAAAAAGATACATATGATTAATGGGA 1011
Db 1565 LysGlnThrLysValLeuSerValSerArgLysLeuAspGlnAlaPheAspLysIleGln 1584
OY 1012 AAGTTACAGAGAAAGATTAAGATTTGAGAGACAGTGTCTCAGATGGAGAGATACAT 1071
Db 1585 AspLeuGlnLysAsnArgLysThrLeuGlnAsnGlnLysAspLeuAlaAsnThrGln 1604
OY 1072 GAGAGCATGAGCAAGAGCTTAAGGACGTGGATTAACACAGCCAGCCAGCCAGAG 1131

```

```

Db 1605 GlyThrLysAspLysAsnValHisGluLeuGlnLysAlaLysArgAlaLeuGluSerGln 1624
OY 1132 CTGGTCGAGCTCCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1191
Db 1625 LeuAlaGlu---LeuLysAlaGlnAsnGlnGlnLysAspLeuGlnLeuThrGln 1643
OY 1192 GAGGTGAGCCGCTGCGAGACCAAGTTA 1218
Db 1644 AspAla---LysLeuArgLeuGlnVal 1651

RESULT 8
A59234
slow myosin heavy chain 3 - quail
C:Species: Coturnix coturnix
C:Date: 19-May-2000 #sequence_rev1sion 19-May-2000 #text_change 08-Sep-2000
C:Accession: A59234
R:Nikoyte Jr., W.; Wang, G.F.; Feldman, J.L.; Miller, J.B.; Wade, R.; Nelson, L.; St
J. Biol. Chem. 271, 17047-17056, 1996
A:title: Isolation and characterization of an avian slow myosin heavy chain gene expr
A:Reference number: A59234; MUID:96291845; PMID:8663323
A:Accession: A59234
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1931 <NIK>
A:Cross-references: GB:053862; NID:g1289513; PIDN:A659912.1; PID:g1289514
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:81-761/Domain: myosin motor domain homology <MMO>

Alignment Scores:
Pred. No.: 1,41e-08 Length: 1931
Score: 253.50 Matches: 123
Percent Similarity: 39.66% Conservative: 86
Best Local Similarity: 23.34% Mismatches: 181
Query Match: 9.28% Indels: 137
DB: 2 Gaps: 20

US-09-502-945-1 (1-1552) x A59234 (1-1931)
OY 19 AGCTAAACTTACTTATGAGAAAGTGTGAATTTGAGGATCCCAATGGAAGTTTGTG 78
Db 1235 LysAlaLysAlaAsnLeuGlnLysMetCysArgSerThrGlnAspGlnMetAsnGlnHis 1254
OY 79 AGGAGCACTTACCTGAATATCAGAGAACTTGTGAACAT----- 117
Db 1255 ArgAsnLysLeuGlnGlnSerGlnArgThrValThrAspLeuSerThrGlnArgAlaLys 1274
OY 118 -----CTTAAAGAGCACTAAGCATTAAGCATTAATTTCTTCG--- 153
Db 1275 LeuGlnThrGluAsnSerGlnLeuSerArgGlnLeuGlnLysGlnAlaPheIleAsn 1294
OY 153 ----- 153
Db 1295 GlnLeuMetArgGlnLysLeuThrTyThrGlnGlnLeuGlnAspLeuLysArgGlnLeu 1314
OY 154 -----GCTGCTAATCTTGAACCGTGTGATGCTGTTGTTGAATGCTGCAGCAT 207
Db 1315 GlnGlnLysAlaLysAlaLysAlaLysAlaLysAlaLysAlaLysAlaLysAlaLys 1333
OY 208 GAAGCTGTTCTTCCCAACCCATTAATGTTTCATATGACAGCCATCGAAGACTGGT 267
Db 1334 AspCysAspLeuLeuArg-----GluGlnTyArgGlnGlnGlnMet 1346
OY 268 AAGAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 321
Db 1347 GlnAlaLysAlaGluLeuGlnArgAlaLeuSerLysAlaAsnSerGlnValAlaGlnIle 1366
OY 322 -----GATAGCGAGCAAGAGAA-----GCAAGTGGCTTAAGAA 354
Db 1367 ArgThrLysTyArgLysThrAspAlaIleGlnArgThrGlnGlnLeuGlnGlnLysLys 1386
OY 355 CAGGTGAACCAAGTTTGAATATCTGAGAGAGCC----- 390

```

Db 1387 LysLeuAlaGlnArgLeuGlnGlnValGlnAlaValAlaAsnAlaLysCys 1406  
QY 391 ---AATTGTAAGAAACCAAGCCTTAATCCAG----- 420  
Db 1407 SerSerLeuGlnLysThrLysHisArgLeuGlnAsnGlnLeuLysPheLeuAlaAsp 1426  
QY 421 -----TGTCAGCCAG 429  
Db 1427 ValGlnArgSerAsnAlaAlaAlaAlaLeuAspLysLysGlnArgAsnPheAspLys 1446  
QY 430 TTGAGGAAGAGCTGAGAGAGCGGAGCGGACCTTGAAAAAGAACTT---GCATCTCAG 486  
Db 1447 IleLeuSerGlnLysThrLysPheGlnLysPheGlnLysSerGlnThrGlnLeuGlnLysSerGln 1466  
QY 487 CAAAGGAAAAAGCGCATTCAGAGAAACATGATGAAA-----AAGCAATA 531  
Db 1467 LysGlnAlaArgSerLeuSerThrGlnLeuPheLysLysAsnAlaTyrGlnLysLeuSer 1486  
QY 532 ACGAAGAAAAAGGAGTACATGCGATCAAAAGATGATGATCTGTCAGAAATATGCCCCA 591  
Db 1487 LeuGlnHisLeuGlnLysThrPheLysArgGlnLysAsnLysAsnLeuGlnGlnLysLeuSerAsp 1506  
QY 592 CTGAGAGCGCCAGGTGAGAAAAAGTTTACAAAGAAAAAGATTTCAGCTAATTAATCACTGAG 651  
Db 1507 LeuThrGlnGlnLeuGlnLysAlaSerGlnLys-----SerLleHisGlnLeuGln 1522  
QY 652 GAAATTCAAAGCCACTGCTCTCTCGGGAATGATGTCACAAAGGTGTGTGGAATAATG 711  
Db 1533 LysValArgLysGlnLeuAspAlaGlnLysLeuGlnLeuGlnAlaLeuGlnLysAla 1542  
QY 712 CGCTATCAGCTGATATAAAACCAACATGAGAGATGAGAGAGAGAGAGAGAGAGAG 771  
Db 1543 GlnAlaSerLeuGlnHisGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 1562  
QY 772 TTCAGAGCAAAAACTAACAGGAGATTTGAATTAAGATCAGGAATAAGAGAA----- 825  
Db 1563 ValLysAlaGlnLysGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 1582  
QY 826 -----TTGGAATGAGACTGATGAAAGCAAAACAACTTGGACAGCAGCAGCAG 876  
Db 1583 ArgAsnHisLeuArgVal---ValAspSerLeuGlnThrSerLeuAspAlaGlnLysThrArg 1601  
QY 877 AAGCAGCCCTGCGAG 924  
Db 1602 -----SerArgAsnGlnAlaLeuArgLeuLysLysLysLysLysLysLysLysLys 1617  
QY 925 GCGCAATGTGAGCACCACATG---CACCTCAGCAGATCTGAAATAGCTCACTCAGTCAA 981  
Db 1618 AsnGlnMetGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1637  
QY 982 GAAAAA-----AGGTATACATATGATTAATTTGGAAAGCTTACAGAGA 1023  
Db 1638 ValLysAlaLeuGlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 1657  
QY 1024 AGAAATGAAGATTTGAGAGACAGTGT----- 1050  
Db 1658 AlaAsnGlnAspLeuLysGlnAsnLysLysLysLysLysLysLysLysLysLysLysLys 1677  
QY 1051 -----GTCCAGCATGAGAGATACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1104  
Db 1678 SerGlnLeuGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1697  
QY 1105 AAGCAGCAGCAGGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1161  
Db 1698 GlnGlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1717  
QY 1162 CTTCCTCTGAGAGCAGAGCCTCTCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1221  
Db 1718 LeuIleAsnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1737  
QY 1222 AGCATGCCACATCTGATTGC 1242  
Db 1738 GlnAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1743

RESULT 9  
152300  
gi|152300| human  
N:Alternate names: gcp372  
C:Species: Homo sapiens (man)  
C:Date: 01-Nov-1996 #sequence\_revision 01-Nov-1996 #text\_change 20-Aug-1999  
C:Accession: 152300  
R:Source: M.; Mitsuoka, Y.; Fujikawa, T.; Nishio, M.; Ikehara, Y.  
Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994  
A:Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized  
A:Reference number: 152300; MUID:95100974; PMID:7802676  
A:Accession: 152300  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-3225 <RES>  
A:Cross-references: GB:D25542; NID:g662389; PIDN:BA05025.1; PID:g808869  
C:Superfamily: glnctln  
  
Alignment Scores:  
Pred. No.: 1.4e-08 Length: 3225  
Score: 253.50 Matches: 119  
Percent Similarity: 40.33% Conservative: 104  
Best Local Similarity: 21.52% Mismatches: 179  
Query Match: 9.28% Indels: 151  
DB: 2 Gaps: 24  
  
US-09-502-945-1 (1-1552) x 152300 (1-3225)  
QY 1 CTTCGTGATGCATCCGAGAGCTAAACTTACTAT-----GAG 39  
Db 2170 ValIleAspGlnAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2189  
QY 40 GAAAGGTGTGAATTTGAGAGATCCCAATTGAAGTTTGGAGAGAGAGAGAGAGAGAGAT 99  
Db 2190 GlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2209  
QY 100 CAGAGACTTGTGAGAGAGCTTAA-----GAGCAACATTAACAT----- 138  
Db 2210 SerLleHisMetGlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 2229  
QY 139 -----AAGAAATTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180  
Db 2230 TrpLysSerLysAlaGlnThrGlnValGlnLeuGlnGlnLysValLysAspThrLeuGln 2249  
QY 181 GGTCTTTGTGAAATGTGCTCAGCATGAGAGCTGTTCTTCCCAACCCATCACTAATGTT 240  
Db 2250 Gly-----GlnAsnLysGlnLeuLeuSerGlnLeuGlnLysLysLysLysLys 2263  
QY 241 CATATG-----CAGACATCGAAACATGCTGTTAAAGAAAGAGATGACTTG 285  
Db 2264 HisLeuTyrHisSerSerGlnAsnGlnLeuAlaLysLysLysLysLysLysLysLys 2283  
QY 286 ATGTCGCACTAGTTTCGTAAGAGAGAGCTTG----- 318  
Db 2284 LysAspGlnLeuThrAspLeuSerLysSerLysLysLysLysLysLysLysLysLysLys 2303  
QY 319 -----GCAGATACCGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 351  
Db 2304 LeuGlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2322  
QY 352 GAGACAGTGAAGAAACATTTTGCATATCT-----GAGGAA 387  
Db 2323 GlnGlnLeuGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2342  
QY 388 GCCAATTTTGAAGAAACCAAGCCTTA----- 414  
Db 2343 IleAsnMetLysGlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2362  
QY 415 ATCCAGTGTGACCACTTGAAG 474  
Db 2363 ValAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2382





Db 1566 GluGluLeuLeuArgLeuGlnAlaGluGlnAlaGlnArgLeuArgGlnAlaGluValGln 1585  
QY 709 ATGGCGTATAGCGTGAAT-----AAAACCAACATGAGAGATGAGCGAGAAAG 759  
Db 1586 ArgAlaArgGlnValGlnValAlaLeuGlnThrAlaGlnArgSerAlaGlnAlaGlnLeu 1605  
QY 760 GAGCACAGCA-----GAGTTCAGAGCAAAAACCTAACAGGATCTGGAATTAAGATCAG 813  
Db 1606 GlnSerLysArgAlaSerPheAlaGlnLysThr---AlaGlnLeuGlnArgSerLeuGln 1624  
QY 814 GAATATAGCAAAATTGAGAAATAGAACTGGATGAAAGCAACACCTTGAGACGAGCAG 873  
Db 1625 GluGlnHisValAlaValAlaGlnLeuArgGlnGlnAlaGlnArgAlaGlnGlnGln 1644  
QY 874 CAGAGAGCAGCCTCGGCGAGAGAGAGAGTGCCTGAGCTAACAGAA-----CTGCTG 924  
Db 1645 AlaGlnAlaLeuArgAlaArgGlnGlnAlaGlnArgGlnLeuGlnArgGlnLeuLys 1664  
QY 925 GGGCAATCTGAGCAGCAACTGCACCTCACCAGATCTGAAATAGCTCAA-----CTC 975  
Db 1665 AlaAsnGlnAlaLeuArgLeuArgLeuGlnAlaGlnGlnValAlaGlnGlnLysSerLeu 1684  
QY 976 AGTCAA---GAAAAAGGTATACATATGATTAATTTGGAAAGTTACAGAGAAAGTAA 1032  
Db 1685 AlaGlnAlaGlnAlaGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 1704  
QY 1033 GAATTTGAGAGACAGTGTCTCCAGCATGGAGATCATGACAGATGAGAAAGAGCTA 1092  
Db 1705 LysAlaGlnGlnGlnGlnAlaValAlaArgGlnArgGlnLeuAlaGlnGlnGlnLysGln 1724  
QY 1093 AGCAGCTGATTAAGCACACCCAGCCAGCCAGCAG----- 1131  
Db 1725 ArgGlnLeu-----AlaGlnGlnLysThrAlaGlnGlnArgLeuAlaAlaGlnGln 1741  
QY 1132 CTGTGTCAGCTCCGACAGACAGCAACAGCAGCTCTCTGAGAGAGAGCAGCTGTGGAA 1191  
Db 1742 LeuLeuArgLeuArgAlaGlnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1761  
QY 1192 GAGGTGACCGCTGCGGACCACTTACCAGATGCCACAA 1233  
Db 1762 GluLeuAlaArgLeuGlnArgGlnAlaAlaAlaAlaThrGln 1775  
  
RESULT 12  
B43402  
myosin heavy chain-B, neuronal - chicken  
N:Contains: myosin ATPase (EC 3.6.4.1)  
C:Species: Gallus gallus (chicken)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 19-Apr-2002  
C:Accession: B43402; A43402  
R:Takahashi, M.; Kawamoto, S.; Adelstein, R.S.  
J. Biol. Chem. 267, 17864-17871, 1992  
A:title: Evidence for inserted sequences in the head region of nonmuscle myosin specific  
yosin.  
A:reference number: A43402; MID:92388144; PMID:1355479  
A:Accession: B43402  
A:Molecule type: mRNA  
A:Residues: 1-2007 <TAA>  
A:Cross-references: GB:M93676; NID:g212448; PIDN:AAA48988.1; PID:g212449  
A:Note: sequence extracted from NCBI backbone (NCBIN:112864)  
C:Comment: Alternatively spliced segments 1 and 2 are found exclusively in nonmuscle myo  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolyase; methylated  
F:1-2007/Product: myosin heavy chain-B, neuronal #status predicted <MTN>  
F:1-211,222-631,653-2007/Product: myosin heavy chain-B, nonmuscle #status predicted <MTN>  
F:88-802/Domain: myosin motor domain homology <MMOT>  
F:178-185/Region: nucleotide-binding motif A (P-loop)  
F:212-221/Region: alternatively spliced segment 1 #status experimental  
F:559-593/Region: actin binding #status predicted

F:632-652/Region: alternatively spliced segment 2 #status experimental  
F:692-714/Region: actin binding #status predicted  
F:875-2007/Domain: coiled coil #status predicted <COI>  
F:875-1315/Region: S2  
F:1316-2007/Region: light meromyosin  
F:129/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted  
F:184/Binding site: ATP (Lys) #status predicted  
F:732,742/Active site: Cys #status predicted  
F:1954/Binding site: phosphate (Thr) (covalent) #status predicted  
F:1987/Binding site: phosphate (Ser) (covalent) #status predicted  
  
Alignment Scores:  
Pred. No.: 2,52e-08 Length: 2007  
Score: 249.50 Matches: 110  
Percent Similarity: 46.58 Conservative: 108  
Best Local Similarity: 23.50% Mismatches: 161  
Query Match: 9.13% Indels: 89  
Dels: 1 Gaps: 20  
  
US-09-502-945-1 (1-1552) x B43402 (1-2007)  
QY 1 CTGTGATGATGATCCGAGAGCTTAACTTATGAGGAAAGTGAATTTGAGGAA 60  
Db 1067 TleThrspleuGlnGlnArgLeuLysLysGlnLysThrArgGlnGlnGlnLys 1086  
QY 61 TCCCAATTGAAGTTT-----TTGAGAGAGCAGCTTACGATATATCAG 102  
Db 1087 AlAlsArgLysLeuAspGlnThrThrAspLeuLnsGlnAlaGlnGln 1106  
QY 103 AGAATCTGTGAGATCTTAAAGACAACTTAACATTAATTTCTTGGCTGCT--- 159  
Db 1107 AlaGlnLleGlnGlnLysLleGlnLeuAlaLysLysGlnGlnGlnAlaAla 1126  
QY 160 -----AATCTGTAAACGCTGTGGNGCTT 186  
Db 1127 LeuAlaArgLyspGlnGlnAlaValAlaGlnLysAsnAlaLeuLysVal----- 1143  
QY 187 TGTTTGAATGTCTCAGCATGAGCTGTTCTTCCCAACCATATTAATGTTATATG 246  
Db 1144 -----TleArgGlnLeuGlnAlaGlnAlaGlnGlnGlnGlnGlnGln 1160  
QY 247 CAGACCATCGAAGA-----CTGCTTAAAGAAAGATACTTGAATGCTGCACTAGT 300  
Db 1161 GluLysAlaSerArgAsnLysAlaGlnLysLysArgAspLeuSerGlnGln 1180  
QY 301 TCCGTAGAGAGAGCTTGCAGATACGACAGCAAGCAAGTGGTTATGAA---CAG 357  
Db 1181 AlaLeuLysThrGlnLeuGlnAspThrLeuAspThrThrAlaGlnGlnGlnLeuArg 1200  
QY 358 GTGAAACAGCTTTGCAATATCTGAG-----GAGCCAAATTTGAAAAACCAAG--- 408  
Db 1201 ThrLysArgGlnGlnGlnValAlaGlnLeuLysAlaLleGlnGlnGlnThrLysAsp 1220  
QY 409 -----GCTTATCCAG-----TGTGACGAGTTGAG 435  
Db 1221 HisGlnAlaGlnGlnGlnLleArgGlnArgHisAlaThrAlaLeuGlnGlnSer 1240  
QY 436 AAGGAGCTGAGAGAGCAGCGAGAGCTTGAAGAAAGATTCATCTCAGCAAGAGAA 495  
Db 1241 GluGlnLeuLeu-----GlnAlaLysArgPheLysAlaAsnLeu-----GluLysAsnLys 1257  
QY 496 AGGCGCATGAGAAAGCATGATGAAAGGAAATTAACGAAAGGAGTCAATGGCA 555  
Db 1258 GlnGlnLysLeuLysAsp-----AsnLysGlnLeuAlaCysGlnValLysValLeuGln 1275  
QY 556 TCAGAGTGTGATCTTGTCTCAGATATTTGCCACATGAGGCCAGCGAGGAAAGGT 615  
Db 1276 GlnValLysAlaGlnSerGlnHisLysArgLysAlaGlnAspAlaGlnGlnGln 1295  
QY 616 ACAAGAGAAAGATTTCAGCTATTATCACTGAGAGAAATTAAGCCAGCTGCTTC 675  
Db 1296 Thr-----AlaLysValThrGlnGlnGlnArgLeuArgValGlnLeuAlaGln 1311

```

QY 676 CGGGAATGATGTCACAAAGGTGTGTGGAGAAATCGCTATCAGCTGAATTAACCAAC 735
   :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1312 Lys-----Alasnllysluenglinsngluleuaspasnlalser 1324
QY 736 ATGGAGAAAGATGAGCCGAAAAAGGACACAGAGCTTCACAGCAAAACTPAACAGGGAT 795
   :: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1325 Serleuengluglualaglysluylsnglyleuysphe---Alalysaspalalaser 1343
QY 796 CTGGAATTAAGATCAGGAATGAGAAATTTG----- 828
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1344 LeuqluserglnleuenglinsphtglnleuengluglulthArGlnlyslu 1363
QY 829 -----AGAAATA---GAACGTGATGAAGCAAAACAACCTTGAACAGAGCAG 873
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1364 AsnleuserSerArglllearglnleuengluglulysasanaenleuengluglnln 1383
QY 874 CAGAAAGCAGCCCTGGCCAGA-----GAGGAGTCCCTGAGACTPAACAGACTG 921
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1384 GluglulguglulglualalarglylsasnleuenglulysglnMetleualaleuqlnla 1403
QY 922 CTGGGCGAATCTGAGCACCACTGCACCTCACAGATCTGAATAGCTCACTCAGTCAA 981
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1404 Leualagluallalyslyslvalaspaspaspleuengllytrllleugllyleuqlnlu 1423
QY 982 GAAAAAAGCTATACATATGATTAATTTGGAAAGTTACAGAGAGAATGAAGAAATTGGAG 1041
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1424 AsnlysllyslleuqlleuqllyAspmetgluserleuserglnarg-----Leuqln 1440
QY 1042 GAACAGTGTCTCCAGATGGGAGACTACATGAGACGATGAACAAAGCTTAAGCAGCTG 1101
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1441 GlulysalameAlaTyAspIlyslleuqlulysThrlyAsnArgleuqlnleuql 1460
QY 1102 GAT-----AGGCACGCCAGGCCACAGCCAGCCAGCTGGTGCAGCTC 1143
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1461 AspsapleuMetValaspIeulaspIhscIlnArgIlnlevalSerasnleuqlulys 1480
QY 1144 CTCAGCAAGCAGAACCACTTCTCTCTGAGAGAGCAGACCTGTCTG-----GAA 1191
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1481 GlnlysllyspheaspIlnleuqlaleglululysasnIleaserAlaArgTyAlaqln 1500
QY 1192 GAGGTGACCGCGCTGGGAGCCAG 1215
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1501 GluArgspARgAlaqlualaglu 1508

RESULT 13
A39638
Plectin - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: A39638; S21876
R:Wiche, G.; Becker, B.; Huber, K.; Welter, G.; Castanon, M.J.; Hauptmann, R.; Stratowe
J. Cell Biol. 114, 83-99, 1991
A:Title: Cloning and sequencing of rat plectin indicates a 466-kD polypeptide chain with
A:Reference number: A39638; MUID:91268156; PMID:2050743
A:Accession: A39638
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-4687 <RIC>
A:Cross-references: EMBL:X59601; NID:91292885; PIDN:CAA42169.1; PID:91561642
C:Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein S
C:Keywords: cytoskeleton; transmembrane protein
F:6-103/Domain: ribosomal protein S10 homology <RS10>
F:184-399/Domain: alpha-actinin actin-binding domain homology <ACT>

```

## Alignment Scores:

```

Pred. No.: 2.89e-08 Length: 4687
Score: 248.50 Matches: 113
Percent Similarity: 44.32% Conservative: 82
Best Local Similarity: 25.68% Mismatches: 164
Query Match: 9.10% Indels: 81
DB: 1 Gaps: 19

```

US-09-502-945-1 (1-1552) x A39638 (1-4687)

```

QY 4 CTGATGCATCGAGACGCTAAAACTTACTTATGAGAAAG-----TGCGAA 51
   ||| ||| : : : : : ||||| ||||| : : : : :
Db 1675 LeuqlnAlaleuaspIleuqlleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqln 1694
QY 52 ATTGAGAAATCCCAATTGAAGTTTGTGAGGAACGACTTGAATATCAGAACTTGT 111
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1695 AlagluAlagluArgAlaArgGlnValGlnValAlaleuqlnleuqlnleuqlnleuqlnleuqln 1714
QY 112 GAAGATCTTAAGACCACTAAGCATTAAGCAATTTCTTGGCTGCTTAATCTTGTAC 171
   ||| : : : : : ||| ||| : : : : :
Db 1715 Glu---ValGluLeuqlnSerlysarproserpe----- 1725
QY 172 CGTGTGTGTGCTTGTGTTGAATGTCTGACGATGAAAGCTGTCTTCCCAACCAT 231
   ||| : : : : : ||||| ||| ||| : : : : :
Db 1726 -----AlagluLystrAlaglnleuqlnleuqlnleuqlnleuqlnleuqlnleuqln 1740
QY 232 ACTAATGTCATPANG-----CAGACCATGAAACACTGTTAAGAAAGAGATGACTTG 285
   ||| : : : : : ||||| : : : : : : : : : : : : : : : : :
Db 1741 ValThrValThrGlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqln 1760
QY 286 ATGCTGCACTAGTTTCCGTTAAGAGCAGCTTGGCAGATACGCAAGAAAGAGCAAGT 345
   ||| : : : : : ||| : : : : : : : : : : : : : : : : :
Db 1761 GluArgAlaArgGluqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqln 1777
QY 346 GCTTATGAACAGGTGAACAAGTTTTCGAAATATCTGAGAAACCAATTTGAAAAAC 405
   ||| ||| : : : : : ||||| : : : : : : : : : : : : : : : : :
Db 1778 AlasngluAlaleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqln 1796
QY 406 AAGCTTTAATTCAGTGTGACCACTTGAAGAGAGCTGAGAGCAGCGAGCAGCTT 465
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1797 LeuqlnAlaleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqln 1816
QY 466 GAAAAAGACTTGCATCTCAGCAGAGAAAAGGCCATTTGAAAAGACATATGAAAAG 525
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1817 Glylys-----Alagluqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqln 1832
QY 526 GAAATTAACGAAAGAAAG-----GAGTACATGGATGCAAGATGTGCTGTGCT 576
   ||||| : : : : : ||||| : : : : : : : : : : : : : : : : :
Db 1833 GluleuqlnlyslGlnArgGlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqln 1852
QY 577 CAGATATTTGCCCACTGAGAGCCAGGTGGA----- 609
   ||| : : : : : ||||| : : : : : : : : : : : : : : : : :
Db 1853 Glnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqln 1872
QY 610 -----AAGTTACAAAGAAAGATTTAGCTATTATCAACTGGAGCA 654
   ||| : : : : : ||| : : : : : : : : : : : : : : : : :
Db 1873 GluglulguleuAlaArgleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqln 1892
QY 655 ATTCAAGCCAGCGTGGCTTCGGGAATGATGTCACAAAGGTGTGGAAGAAATGGCG 714
   ||||| : : : : : ||||| : : : : : : : : : : : : : : : : :
Db 1893 LeuqlnAlagluLeuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqln 1905
QY 715 TATCAGCTGAATTAACCAACATGAGAGATGAGAGAGAGAGAGAGAGAGAGAGTTC 774
   ||| : : : : : ||| : : : : : : : : : : : : : : : : :
Db 1906 Valleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqln 1921
QY 775 AGACCAAAACTAACAGAGGATCTTGAATTA-----GATCAG 813
   ||||| : : : : : ||||| : : : : : : : : : : : : : : : : :
Db 1922 SergluylsSerlyslGlnArgleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqln 1941
QY 814 GAAATTAAGAAATTAAGAAATGAAGCTGATGAAGCAAAAGCAACTTGAACAGAGAG 873
   ||| : : : : : ||||| : : : : : : : : : : : : : : : : :
Db 1942 GluAlaAlaArgleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqln 1961
QY 874 CAGAAAGCAGCCCTGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 930
   :: : : : : : ||| ||| : : : : : ||||| ||||| : : : : :
Db 1962 Gluspalalaleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqln 1981
QY 931 TCTGAGCAGCACTGCACTG-----ACAGATCTGAATAGCTCAACTCAGTCAAGAAAAA 987
   ||| ||| : : : : : ||||| : : : : : : : : : : : : : : : : :
Db 1982 IleSerluAlaThrArgleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqlnleuqln 1999

```





Db 1466 LysLysPheAspGlnLeuLeuAlaIcLysAsnLleSerAlaArgHisAlaGlu 1485

QY 1039 GAGGAAACAGTGTGTCACGATGGGAGAGTACATGACGATGAAAGAAAGCTA--AGG 1095

Db 1486 ArgAspArgGlnAlaGlnAlaAspAlaArgGluLysGluThrLysAlaLeuSerLeuAlaArg 1505

QY 1096 CAGCTGATTAAGCACAGCAGCGCCACAGCCAGCCAGCTGGGACGCTCTCAGCAAGAG 1155

Db 1506 AlaLeuAspGlnAlaLeuGlu-----AlaGlnAspGluPheGluArgLeuAsnLysGln 1523

QY 1156 -----AAC 1158

Db 1524 LeuArgAlaGluMetGluAspLeuMetSerSerLysAspAspValGlyLysAsnValHis 1543

QY 1159 CAGCTTTCCTCGGACAGGACGCTCTCGGAAGAGGTGGACCGGCTCGGACCCAGTTA 1218

Db 1544 GluLeuGluLysSerLysArgAlaLeuAspGlnValGluGluMetArgThrGlnLeu 1563

QY 1219 CCCAGCATG 1227

Db 1564 GluGluLeu 1566

## RESULT 15

myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)  
C:Species: Dictyostelium discoideum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 08-Sep-2000  
C:Accession: T18278; T30579  
R:Hammer III, J.A.; Jung, G.  
J. Biol. Chem. 271, 7120-7127, 1996  
A:Title: The sequence of the dictyostelium myo I heavy chain gene predicts a novel, dimeric myosin  
A:Reference number: 218854; MUID:96215148; PMID:8636147  
A:Accession: T18278  
A:Status: preliminary; translated from GR/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 1-2245 <HAM>  
A:Cross-references: EMBL:U42409; NID:g1150765; PID:g1150766; PIDN:AAA85186.1  
R:Titus, M.A.; Kuspa, A.; Loomis, W.F.  
Proc. Natl. Acad. Sci. U.S.A. 91, 9446-9450, 1994  
A:Title: Discovery of myosin genes by physical mapping in Dictyostelium.  
A:Reference number: 220873; MUID:95023928; PMID:793787  
A:Accession: T30579  
A:Status: preliminary; translated from GR/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 1-190, 'F', 192-283, 'T', 285-290, 'R', 292-331, 'IE', 'WM', 337-338, 'LK', 342, 'YRMKS  
A:Cross-references: EMBL:L35322; NID:g1039360; PID:g1039361; PIDN:AAA79858.1  
C:Genetics:  
A:Gene: myoJ  
A:Introns: 257/3; 307/3  
C:Superfamily: myosin heavy chain; myosin motor domain homology <MMO>  
C:Keywords: ATP; P-loop  
R:84-809/Domain: myosin motor domain homology <MMO>

### Alignment Scores:

Pred. No.:	4,836-08	length:	2265
Score:	245.00	Matches:	108
Percent Similarity:	37.94%	Conservative:	87
Best Local Similarity:	21.01%	Mismatches:	195
Query Match:	8.97%	Indels:	124
DB:	2	Gaps:	12

US-09-502-945-1 (1-1552) x T18278 (1-2245)

```

0Y      4 CTGATGATCCGGAAGCTAAACTTACTTATGAGCAAAAGTGTGAATTGAGCAATCC 63
          ||||| ||| ||||| |||||
Db      1064 LeuApHisSerLysLeuAnLysLysLeuGlnLysAspLeuSerAspLinhisAp 1083
          ||||| ||||| ||||| |||||
0Y      64 CAATTGAGTTTGGAGCAAGCACTTAGCGTAATTCAGAGCACTTGTAAGATCTTAA 123
          ||||| ||||| ||||| |||||
Db      1084 SerIecLysLeuGlnSerGlnPheasnGluThGluGlnLeuIecInlPheLys 1103
          ||||| ||||| ||||| |||||
0Y      124 GAGCACTAAAGCATTAAGAAATTTCTTGCGTGT- 159
          ||||| ||||| |||||

```

Dd	1104	GlnGlnSerGlnGluLeuSerSerLysLeuSerLysThrThrGlnGlnLeuAspPheAsn	11233
Qy	160	-----AAACTTGTGAACGGTGT	177
Dd	1124	LysGlnGluPheAspArgLeuSerGlnGluArgAspThrAspAsnThrAsnAsnGlnLeu	1143
Qy	178	GGTGGCTCTTGTGTAATGTGCT-----CAGCATGAAAGCTGTCTTCTTC	222
Dd	1144	GluIleGlnGlnLeuLysLysAlaAsnSerThrLeuGlnGluAspTyrPheSerLeuSer	1163
Qy	223	CAAAACCATACTATGTCTTCATTGCAAGACCATGGAA-----AGACTGGT	267
Dd	1164	GlyIleArgAspAsnLeuGluArgGlnValLeuGlnLeuArgAspGluAsnGlnIle	1183
Qy	268	AAAGAAGAGAGTGAAGCTTGATGTCTGCACATGTATCCGTAAAGACAGCTTGACA-----	321
Dd	1184	LysGluArgLeuAspSerLeuGlyGlnGlnSerSerGlnPheGlnSerGlyAlaAlaLeu	1203
Qy	322	GATACGACGACAAAGAGAGCAAGTGTCTTATGAACAGGTGAACAGTTTGTGCAAAATATCT	381
Dd	1204	GluLysGlnGlnLeuGlnGlnLeuValGlnGlnGlnSerGlnGlnLeuIleLysLeuSer	12233
Qy	382	GAGGAAGCCAAATTTTGAAAAACCAAGCCTTATATCAGTGTGACCATGTGAGCAAGAG	441
Dd	1224	SerGlnLysLeuGlySerGlnGlnGlnAlaLysLysGlnIleAsnGlnLeuGlnLeuGln	1243
Qy	442	CTGGAGAGCGAGCGGAGCCACTTGAAGAAAGACTTGCTACACCAAGAGAAAGGCC	501
Dd	1244	LeuThrAspHisLysSerLysLeuGlnIleGlnLeu-----GlnLeuThrGlnLeuSer	1261
Qy	502	ATTGAGAAGACATGATGAAGAAAGAAATPACGAAGAAAGAGGATACATGGATCAAG	561
Dd	1262	AsnGluLys-----IleLysLysLeuLysGlyLysLeuGlnGlnIleTyrGlnAspGluLys	1279
Qy	562	ATGTGATCTTGTCTCAGAAATATTGCCCACTGAGAGCCCAAGTGAAGAAAGTTACAAG	621
Dd	1280	-----LysGlnGlnGlnGlnLeuGlnGlnGlnGlnGln	1291
Qy	622	GAAGAAGATTCAGCTATTATCAACTGGAGGAATTCAGAGCAGCTGGCTTCGGGAA	681
Dd	1292	SerLysGlnSerValGluAspGlnLysAsnSerLeuIleThrGlnLeuThrThrValLys	1311
Qy	682	ATGATGATCACAAGGTGTGTGGAGAAATGCGCTATCAAGCTCAATAAACCAATCGAG	741
Dd	1312	PheGlnSerThrGlnValSerThrAsnValSerHisGlnLysGlnLysIleThrThrLeu	1331
Qy	742	AAGGATGAGCAGAAAAGAGCAGACAGAGTTCAGAGCAAAACTACAGGAGATCTTGA	801
Dd	1332	LysSerThrIleGlnGlnLeuAsnLysSerIle---GlyLysLeuGlnAlaGlnGlnLys	1350
Qy	802	ATTAAAGATCAGAAATPAGAAATTGAGATATGAATCGATGATGAAGCAACACAC---	858
Dd	1351	AsnLysAspAspGlnIleArgLysIleGlnPheGlnLeuAsnAspGlnLysGlnPhe	1370
Qy	858	-----	858
Dd	1371	ThrArgGlnThrLysGlnPheSerAspLeuGlnSerGlnGlnSerIleAspArgProLys	1390
Qy	859	-----TTGGA	864
Dd	1391	SerGluIleThrIleHisSerLeuGluArgThrAsnGlnIleThrLeuLysSerAspPheGln	1410
Qy	865	CAGAGCAGCAGCAAGCAGCCCTGCCACAGAGAGATCC-----	903
Dd	1411	ArgValGlnGlnSerLeuLysGlnGlnArgAspCysGlnGlnIleTyrLysAspThrIle	1430
Qy	904	-----CTGACACTPACAGAGATCTCTGGCGGAATTCGAGCAGCA	942
Dd	1431	AsnArgLeuGluAsnGlnValLysGlnLeuThrGlnLeuLysGlnArgPheGluAsnGln	1450
Qy	943	CTGACACTCACACAGATCTGAATAGCTCAACTGACAAAGAAAGATATCATATGAT	1002
Dd	1451	PhePheValAlaLysGlnGlnAsnSerAsnGlnIleThrGlnGlnSerValTyrLeuLysGln	1470

```
OY 1003 A A A T T G G A A G T T A C A G A G A A A T G A A A T T G G A G G A C A G T G T C C A G C A T G G 1062
      :::::|||||
Db 1471 V a l T h r T h r G l n m e t g l n g l n a s n g l n s e r A r g I l e g l u a r g l u l e u ----- 1486
OY 1063 A G A G T A C A T G A G A C G A T G A A G C A A A G G C T A A G C A C T G G A T A A G C A C A G C C A G C C A C A 1122
      ||| ||||| ::::: |||
Db 1487 -----G l n g l u l y s l y s g l n h i s l e t h r A r g l l e a s p a s p g l u a r g a s p g l u l e u 1503
OY 1123 G C C C A G C A G C T G G T G C A G C T C C T C A G C A A C C A G C A C C A G C T T C C T G A G A G G C A G A G C 1182
      :::::||||| ||||| ::::: |||
Db 1504 L y s l y s g l n l e u t h r g l n l e u g l n g l n h i s g l u c l n s e r S e r t h r g l n l e u l e u 1523
OY 1183 C T G T C G A A G A G G T G A C C G G T G C G G A C C C A G T T A C C A G A T G C C A C A A T C T G A T T G C 1242
      ||::: ||||| |||
Db 1524 A l a g l n a s n g l u l e u g l u a r g l e u A r g l y s ----- 1533
OY 1243 T G A C C T G A T G A C A G A G T G A A T A A T G A T T A C A A A G 1284
      ||| ::::: ||||| |||
Db 1534 -----L y s g l u l e u l y s t y l y s g l u 1540
```

Search completed: March 21, 2003, 13:04:27  
Job time : 70.6459 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 21, 2003, 12:48:47 ; Search time 39.5926 Seconds  
(without alignments)  
10292.707 Million cell updates/sec

Title: US-09-502-945-1  
Perfect score: 2732  
Sequence: 1 ctctgcatgcacccgagaa.....aaatgaacttttaagaaga 1552

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 609376 seqs, 131287123 residues  
Total number of hits satisfying chosen parameters: 1218752

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ n2p model -DEV=xlp  
-Q=/cgn2\_1/USPRO/spool/US09502945/runat\_14032003-101100\_19241/app\_query.fasta.1.10979  
-DB=pendig\_patents\_AA\_New -QFMT=fastan -SUFFIX=rapn -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09502945.ecgn.1.1.226 @runat.14032003.101100.19241 -NCPu=6 -ICpu=3  
-NO\_XLPRY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Pending Patents AA\_New:\*  
1: /cgn2\_6/ptodata/2/paa/PCF\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	666	24.4	264	7	US-60-443-566-4443 Sequence 4443, Ap
2	666	24.4	362	7	US-60-443-566-4444 Sequence 4444, Ap
3	537	19.7	172	6	US-10-276-774-1921 Sequence 11291, Ap
4	280	10.2	1294	6	US-10-282-1224-61292 Sequence 68912, A
5	268.5	9.8	3852	5	US-09-724-676-68911 Sequence 68911, A
6	268.5	9.8	3852	5	US-09-724-676-68912 Sequence 68912, A
7	268.5	9.8	3852	5	US-09-724-676-68913 Sequence 68913, A
8	268.5	9.8	3852	5	US-09-724-676-68911 Sequence 68911, A
9	268.5	9.8	3852	5	US-09-724-676-68912 Sequence 68912, A
10	268.5	9.8	3852	5	US-09-724-676-68913 Sequence 68913, A
11	268.5	9.8	3864	5	US-09-724-676-68923 Sequence 68923, A

12	268.5	9.8	3864	5	US-09-724-676-68924	Sequence 68924, A
13	268.5	9.8	3864	5	US-09-724-676-68925	Sequence 68925, A
14	268.5	9.8	3864	5	US-09-724-676-68923	Sequence 68923, A
15	268.5	9.8	3864	5	US-09-724-676-68924	Sequence 68924, A
16	268.5	9.8	3864	5	US-09-724-676-68925	Sequence 68925, A
17	268.5	9.8	3868	5	US-09-724-676-68914	Sequence 68914, A
18	268.5	9.8	3868	5	US-09-724-676-68914	Sequence 68914, A
19	268.5	9.8	3873	5	US-09-724-676-68926	Sequence 68926, A
20	268.5	9.8	3873	5	US-09-724-676-68926	Sequence 68926, A
21	268.5	9.8	3898	5	US-09-724-676-68931	Sequence 68931, A
22	268.5	9.8	3898	5	US-09-724-676-68931	Sequence 68931, A
23	268.5	9.8	3910	5	US-09-724-676-68917	Sequence 68917, A
24	268.5	9.8	3910	5	US-09-724-676-68917	Sequence 68917, A
25	264	9.7	3861	5	US-09-724-676-68907	Sequence 68907, A
26	264	9.7	3861	5	US-09-724-676-68908	Sequence 68908, A
27	264	9.7	3861	5	US-09-724-676-68909	Sequence 68909, A
28	264	9.7	3861	5	US-09-724-676-68907	Sequence 68907, A
29	264	9.7	3861	5	US-09-724-676-68908	Sequence 68908, A
30	264	9.7	3861	5	US-09-724-676-68909	Sequence 68909, A
31	264	9.7	3873	5	US-09-724-676-68919	Sequence 68919, A
32	264	9.7	3873	5	US-09-724-676-68920	Sequence 68920, A
33	264	9.7	3873	5	US-09-724-676-68921	Sequence 68921, A
34	264	9.7	3873	5	US-09-724-676-68919	Sequence 68919, A
35	264	9.7	3873	5	US-09-724-676-68920	Sequence 68920, A
36	264	9.7	3873	5	US-09-724-676-68921	Sequence 68921, A
37	264	9.7	3877	5	US-09-724-676-68910	Sequence 68910, A
38	264	9.7	3877	5	US-09-724-676-68910	Sequence 68910, A
39	264	9.7	3889	5	US-09-724-676-68922	Sequence 68922, A
40	264	9.7	3889	5	US-09-724-676-68922	Sequence 68922, A
41	264	9.7	3907	5	US-09-724-676-68928	Sequence 68928, A
42	264	9.7	3907	5	US-09-724-676-68928	Sequence 68928, A
43	264	9.7	3919	5	US-09-724-676-68916	Sequence 68916, A
44	264	9.7	3919	5	US-09-724-676-68916	Sequence 68916, A
45	262	9.6	860	6	US-10-072-012-838	Sequence 838, App

## ALIGNMENTS

```

RESULT 1
US-60-443-566-4443
: Sequence 4443, Application US/60443566
: GENERAL INFORMATION:
: APPLICANT: CARCILL, Michele
: APPLICANT: BEGOVICH, Ann
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: C1001447
: CURRENT FILING DATE: 2003-01-30
: NUMBER OF SEQ ID NOS: 25102
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 4443
: LENGTH: 264
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-443-566-4443

Alignment Scores:
Pred. No.: 3.28e-43 Length: 264
Score: 666.00 Matches: 132
Percent Similarity: 100.00% Conservative: 0
Best local Similarity: 100.00% Mismatches: 0
Query Match: 24.38% Indels: 0
DB: Gaps: 0

US-09-502-945-1 (1-1552) x US-60-443-566-4443 (1-264)
QY 16 GAGAGCACTAAGCTTATGAGCAAGCTGCAATGGAATCCATGCAAGTT 75
DB 128 GAGAGCACTAAGCTTATGAGCAAGCTGCAATGGAATCCATGCAAGTT 147
QY 76 TTGAGCACTAAGCTTATGAGCAAGCTGCAATGGAATCCATGCAAGTT 135

```

DB 148 LeuArgAsnAspLeuAlaGluTyrGlnArgThrCysGluAspLeuLysGlnLeuLys 167  
QY 136 CATTAAGATTCTCTGCGTGTATCTATGTAACCGTGGTGGCTTTGTTGAAA 195  
DB 168 HisLysGlnPheLeuLeuAlaAsnThrCysAsnArgValGlyGlyLeuCysLeuLys 187  
QY 196 TGTGCTCAGCATGAGCTGTCTTCCCAACCCATCAATGTCATATGACAGCATC 255  
DB 188 CysAlaGlnHisGlnAlaValLeuSerGlnThrHisThrAsnValHisMetGlnThrIle 207  
QY 256 GAAAGCTGTTTAAGAAAGATGACTGTGACCTACTGTTCCGTAAGGACGAC 315  
DB 208 GlnArgLeuValLysGlnArgAspAspLeuMetSerAlaLeuValSerValArgSerSer 227  
QY 316 TTGGCGATATACGACGAAGAGAGAGAGAGCTTATGAACAGGTGAACAAGTTTGCAA 375  
DB 228 LeuAlaAspThrGlnGlnArgGlnAlaSerAlaTyrGlnGlnValLysGlnValLeuGln 247  
QY 376 ATATCTGAGGAAGCCAAATTTTGAAAAACCAAGCCT 411  
DB 248 IleSerGlnGlnAlaAsnPheGlnLysThrLysAla 259

## RESULT 2

US-60-443-566-4444  
; Sequence 4444, Application US/60443566  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele  
; APPLICANT: BEGOVICH, Ann  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001447  
; CURRENT APPLICATION NUMBER: US/60/443,566  
; CURRENT FILING DATE: 2003-01-30  
; NUMBER OF SEQ ID NOS: 25102  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4444  
; LENGTH: 362  
; TYPE: PRN  
; ORGANISM: Homo sapiens  
US-60-443-566-4444

## Alignment Scores:

Pred. No.: 3,41e-43 Length: 362  
Score: 666.00 Matches: 132  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 24.38% Indels: 0  
DB: 7 Gaps: 0

US-09-502-945-1 (1-1552) x US-60-443-566-4444 (1-362)

QY 16 GAGAAAGCTTAAATCTTACTTATGAGAAAAGTGTGAATTCAGGAATCCCAATTGAAGTTT 75  
DB 226 GlnLysLeuLysLeuThrLysGlnLysCysGlnLysLeuSerGlnLeuLysPhe 245  
QY 76 TTGAGAACGACTTACCTGAATATTCAGAACTTGAAGATCTTAAAGACAACTTAAG 135  
DB 246 LeuArgAsnAspLeuAlaGlnTyrGlnArgThrCysGluAspLeuLysGlnLeuLys 265  
QY 136 CATTAAGATTCTCTGCGTGTATCTATGTAACCGTGGTGGCTTTGTTGAAA 195  
DB 266 HisLysGlnPheLeuLeuAlaAsnThrCysAsnArgValGlyGlyLeuCysLeuLys 285  
QY 196 TGTGCTCAGCATGAGCTGTCTTCCCAACCCATCAATGTCATATGACAGCATC 255  
DB 286 CysAlaGlnHisGlnAlaValLeuSerGlnThrHisThrAsnValHisMetGlnThrIle 305  
QY 256 GAAAGCTGTTTAAGAAAGATGACTGTGACCTACTGTTCCGTAAGGACGAC 315  
DB 306 GlnArgLeuValLysGlnArgAspAspLeuMetSerAlaLeuValSerValArgSerSer 325  
QY 316 TTGGCGATATACGACGAAGAGAGAGAGAGCTTATGAACAGGTGAACAAGTTTGCAA 375

DB 326 LeuAlaAspThrGlnGlnArgGlnAlaSerAlaTyrGlnGlnValLysGlnValLeuGln 345  
QY 376 ATATCTGAGGAAGCCAAATTTTGAAAAACCAAGCCT 411  
DB 346 IleSerGlnGlnAlaAsnPheGlnLysThrLysAla 357

## RESULT 3

US-10-276-774-1921  
; Sequence 1921, Application US/10276774  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; APPLICANT: Tang, Y, Tom et al  
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides  
; FILE REFERENCE: 21272-030  
; CURRENT APPLICATION NUMBER: US/10/276,774  
; CURRENT FILING DATE: 2002-11-18  
; PRIOR APPLICATION NUMBER: 09/560,875  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 09/496,914  
; PRIOR FILING DATE: 2000-02-03  
; NUMBER OF SEQ ID NOS: 2700  
; SOFTWARE: Custom  
; SEQ ID NO 1921  
; LENGTH: 172  
; TYPE: PRN  
; ORGANISM: Homo sapiens  
US-10-276-774-1921

## Alignment Scores:

Pred. No.: 2.82e-33 Length: 172  
Score: 537.00 Matches: 117  
Percent Similarity: 70.21% Conservative: 15  
Best Local Similarity: 62.23% Mismatches: 40  
Query Match: 19.66% Indels: 16  
DB: 6 Gaps: 3

US-09-502-945-1 (1-1552) x US-10-276-774-1921 (1-172)

QY 670 GCTTCTCGGAAATGATGATGTCACAAAGTGTGTGGAATGCGCTATGAGTAATAA 729  
DB 1 AlaSerArgGlnMetAspValThrLysValCysGlyGlnMetArgTyrGlnLeuAsnLys 20  
QY 730 ACCAATGAGAGAGATGATGAGCAGAAAGAGCAGACAGAGTTCAGAGCAAAACTAAC 789  
DB 21 ThrAsnMetGlnLysAspGlnAlaGlnLysGlnHisArgGlnPheArgAlaLysThrAsn 40  
QY 790 AGGATCTTGAATTTAAGATGTCAGAAATGAGAAATGAGATGAGTGAAGTGAAGC 849  
DB 41 ArgAspLeuGlnLysLysAspGlnGlnLysLeuArgLysLeuAspGlnSer 60  
QY 850 AAACAACTTGGACAGAGCAGCAGAAAGGAGCCCTGGCCAGAGAGAGTGCCTGAGA 909  
DB 61 LysGlnHisLeuGlnGlnGlnGlnGlnLysAlaAlaLeuAlaArgGlnLysLeuArg 80  
QY 910 CTAAAGCAACTGTGGCGCAATGTGAGCAACAATGCACTGACCAAGATCTGAAATAAGCT 969  
DB 81 LeuThrGlnLeuLeuGlyLysSerGlnHisGlnLeuHisLeuThrArgGlnGlnLysAsp 100  
QY 970 CAACCTGATCAAGAAAAGATATACATATGATTAATTTGGAAATCTACAGAGAAACAAT 1029  
DB 101 SerIleGlnGln-----SerPheSerLysGlnAlaLysAlaGlnAlaLeuGln 116  
QY 1030 GAAAGATTGGAGAGACAGTGTGTCCAGCATGGAGATGACATGAGCATGAAGCAAGG 1089  
DB 117 AlaGlnGlnArgGlnGln-----GlnLeuThrGlnLys 127  
QY 1090 CTAAAGCACTGATGATGAGCAGACAGCCAGCCAGCCAGCTGCTGACCTCCTCAGC 1149  
DB 128 IleGlnGlnMetGlnAlaGlnHisAspLysThrGlnLysGlnGlnLysThrLeuLeuThr 147  
QY 1150 AAGCAGAACACACTTCTCTGAGAGAGCAGAGCCTGTGCAAGAGAGTGAACGGCTGGCG 1209  
DB 148 SerGlnAsnThrPheLeuThrLys-----LeuLysGlnGlnLysCysThrLeuAla 164

```

Qy      1210 ACCCAGTTACCCAGACATGCCACAA 1233
          ::|||
Db      165 LysLysLeuGlnGlnIleSerGln 172

RESULT 4
US-10-282-122A-61292
: Sequence 61292, Application US/10282122A
: GENERAL INFORMATION:
: APPLICANT: Wang, Liangsu
: APPLICANT: Zamudio, Carlos
: APPLICANT: Malone, Cheryl
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari
: APPLICANT: Zyskind, Judith
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John
: APPLICANT: Carr, Grant
: APPLICANT: Yamamoto, Robert
: APPLICANT: Forsyth, R.
: APPLICANT: Xu, H.
: TITLE OR INVENTION: Identification of Essential Genes in Microorganisms
: FILE REFERENCE: EILTRA.034A
: CURRENT APPLICATION NUMBER: US/10/282.122A
: CURRENT FILING DATE: 2003-02-20
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/230,335
: PRIOR FILING DATE: 2000-09-06
: PRIOR APPLICATION NUMBER: 60/230,347
: PRIOR FILING DATE: 2000-09-09
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/267,636
: PRIOR FILING DATE: 2001-02-09
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: Remaining Prior Application data removed - See File Wrapper or PALM.
: NUMBER OF SEQ ID NOS: 78614
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 61292
: LENGTH: 1294
: TYPE: PRF
: ORGANISM: Legionella pneumophila
US-10-282-122A-61292

Alignment Scores:
Pred. NO.:      2,47e-13      Length:      1294
Score:          280.00      Matches:      106
Percent Similarity: 47.82%      Conservative: 113
Best Local Similarity: 23.14%      Mismatches:  137
Query Match:      10.25%      Indels:      102
DB:              Gaps:        21
US-09-502-945-1 (1-1552) x US-10-282-122A-61292 (1-1294)

Oy      52 ATTGAGCAATCCCAATGCAATTTTGTAGAGCAAGCATTTACCTGAATTCAGAGAACTTGT 111
          ::||| | ||||| ||||| |||  ::  |||
Db      686 LengGluLeuSerGlnLeuLysLeuGlnLeuGlnGlnGlnGlnLysLeuGlnPheGlnLeu 705
          ::||| | ||||| ||||| |||  ::  |||

112 GAAGATCTTAAGAGCAACTAAACATAAAGAATTTCTTGCGTGGCTGTACTTGTAAAC 171
          ::||| | ||||| ||||| |||  ::  |||
Db      706 LysGluLeuLysGluLysIle--HisGlnIleInThrIleuSer-----Lys 720
          ::||| | ||||| ||||| |||  ::  |||

172 CGGTGTGGTGTCTTTGTTGTAATATGCTCAGCATGAAAGCTGTTCTTCCCAACCCAT 231
          ::||| | ||||| ||||| |||  ::  |||

```

Db	721	ArgLeuGly	-----	GlulLeuGlnThrGlnIlyLys	731
Qy	232	ACTAATGTTCATATG	---CAGACCATCGAAGACCTGGTTAAGAAAGACATCTTGATG	288	
Db	732	ThrAsnThrHisGlnIuGlnIuThrIleGlnArgIleThrLysGlnLysSerLeuLaaSP	751		
Qy	269	TCTGCATGACTTTCCGTAAGACAGCTTGGCAGATACCCACAAAGACAAAGCTCT	348		
Db	752	SerSerLeuGlnIuSerLeuArgLysGlnLeuHisGlnLeuAlaLysGlnLysGlnArgSerLeu	771		
Qy	349	TAT	-----GAACAGGTGAAACAAAGTTTCCAAATATCTGAG	384	
Db	772	HisLysThrLeuGlnIuGlnLysGlnLeuGlnValGlnGlnLeuGlnIuGlnLeuAlaLys	791		
Qy	385	-----	GAAGCCAAATTTGAAAAAAC	405	
Db	792	LysGlnLysGlnLysAsnLeuAlaLeuLysLysAlaAspLysGlnSerGlnHisGlnLysSer	811		
Qy	406	-----AAGCTTTAATC	-----CAGTGGACCACTTGAGCAAGCACTG	444	
Db	812	LeuAspLysSerAlaIleGlnIuSerLeuThrSerGlnLeuAsnGlnLeuLysLeuGlnIu	831		
Qy	445	GAGAGGCAAGCGCGAGCACTTGAA	-----AAAGAACTGTGCATCTCAGCAAGAGAA	495	
Db	832	GlnLysGln	---GlnThrLeuGlnLeuGlnLeuLysSerLeuArgLysGlnIleGlnIu	850	
Qy	496	AGGCGCATGTGAAGACATGATGATGAAAGAAAGAAATACCAAGAAAGAGAGTACATCGCA	555		
Db	851	GlnThrLeuValIaIaGlnIuGlnLysGlnIuGlnLysGlnIuGlnLysGlnIuLys	867		
Qy	556	TCAAAGATGTTGATCTTGTCAGAAATATTTCCCAACTGAGAGCCAGCGTGAAGAAAGTT	615		
Db	868	-----	SerAsnThrHisGln	879	
Qy	616	ACAAAGGAAAGATTTACGCTATTATCACTGAGAGAAATTCAAAGCCAGCTG	669		
Db	880	ThrLysGlnLysSerLeuAlaAspSerAlaLeuGlnIuSerLeuArgLysGlnIuLysGlnIu	899		
Qy	670	-----GCTTCTCGGGAATGGATGTGCACAAAGCTGTGAGGAATTCGCG	714		
Db	900	LeuThrArgLysAsnGlnIuGlnLysGlnIuLysLeuThrLysGlnValHisSerLeuSer	919		
Qy	715	TATCAGCTGATATAAACCAACATG	-----GAAGAGATGAGCAGCAAG	759	
Db	920	GlnIuGlnLeuGlnIuLysGlnLeuGlnIleArgGlnPheGlnLysGlnLeuGlnIuLys	939		
Qy	760	GAGCAGCAGA	-----GAGTTCAGACGCAAAACTTAACAGGAGATCTTGAATTT	804	
Db	940	GlnLysArgValIaGlnIuSerGlnLysGlnLysAlaSerAlaLysArgThrValAlaSer	959		
Qy	805	AAAGCATCGAAGATFAGAAATTGACATATGAAATGAAATGCAAGCAACCAACACTGGAA	864		
Db	960	LeuArgGlnIuValIaSerAsnLeuLysLeuGlnIuGlnIleGlnLysGlnIuValIaIleGln	979		
Qy	865	CAGGAGCAGCAGAAAGCAGCCCTGGCCAGAGAG	-----GAGTGGCTGAGACTAACA	915	
Db	980	GlnLysGlnLysGlnLysSerSerLeuIleSerGlnGlnIuSerLysGlnIleIleAlaLeuGlnIu	999		
Qy	916	GAACTGCTGGCGAATCTGAGCAGCAACTGCACTCACCAGATCTGGAATTAAGCTCAACTG	975		
Db	1000	GlnIleIleGlnIaAspGlnLysArgGlnLeuGlnIuLysIleLysIleGlnIuLys	1019		
Qy	976	AGTCAGCAAAAAAGCTTACATATGATTAATTCGGAAATTTACACAGAAAGAAATGAGAA	1035		
Db	1020	ValSerAlaAsnGln	-----GlnLeuGlnLys	1032	
Qy	1036	TTGAGAGAACAGTGTCTCCAGCATGGAGAGTACATGACATGAAACCAAGGCTAAG	1055		
Db	1033	LeuSerLysGlnLysAsnLeuHisAsnLysAsnThrValGlnLysPheLysLysLysLeuAsn	1052		
Qy	1096	CAGCTGGAT	-----AAAGCAGGCCAGGCCAGCCAGCAGCTGGTG	1137	







APPLICANT: CompuGen LTD  
TITLE OF INVENTION: Variants of alternative splicing  
FILE REFERENCE: 129181.4 CompuGen  
CURRENT APPLICATION NUMBER: US/09/724,676  
CURRENT FILING DATE: 2000-11-28  
NUMBER OF SEQ ID NOS: 97222  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 68913  
LENGTH: 3852  
TYPE: PRF  
ORGANISM: Homo sapiens  
US-09-724-676-68913

Alignment Scores:  
Pred. No.: 2,18e-12 Length: 3852  
Score: 268.50 Matches: 121  
Percent Similarity: 41.23% Conservative: 107  
Best Local Similarity: 21.88% Mismatches: 162  
Query Match: 9.83% Indels: 163  
DB: 5 Gaps: 22

US-09-502-945-1 (1-1552) x US-09-724-676-68913 (1-3852)

OY 1 CTCTGATGATCCGAGAGCTTAACTTATGAGAAAAGTGTGA--ATTGAG 57  
DB 1797 MetLeuAsnIleSerIleArgLeuGlnAlaValGluLysLeuGlnAlaIleSer 1816  
OY 58 GAA-----TCCCAATTGAATTTTGAAGAACGAC 87  
DB 1817 GluThrSerSerGlnLeuGlnHisAlaValThrGlnThrGlnLeuMetArgGluSer 1836  
OY 88 TTAGCTGATATATACAGAACTTGTGAAGACTTAAAGACAACTAAAGATTT 147  
DB 1837 PheArgGlnLysGlnGlnAlaThrGlnSerLysGlnGlnGluLeuArgGluArg 1856  
OY 148 CTCTGGCTCTTAATCTTAAACCGTGTGGTCTGTGTAATGTCGACGAT 207  
DB 1857 LeuHisGluGluSerArgAlaArgGlu-----GlnLeuAlaValGluLeuSerLysAla 1874  
OY 208 GAAGCTGCTTTCCCAACCCATCTAATGTCATGACAGACCACTGAAGACTGGT 267  
DB 1875 GluGlyValIleAspGlyTyrAlaAspGluLysThrLeu-----PheGluArgGlnIle 1892  
OY 268 AAGAAAGAGATGACTTGATGTCTGCATA-----GTTCCGCTAAG 309  
DB 1893 GlnGluLysThrAspIleIleAspArgLeuGlnGlnLeuLeuLysAlaSerAsnArg 1912  
OY 310 AGAGCTTGGCAGATACGCAAGCA----- 336  
DB 1913 LeuGlnGlnLeuGlnAlaGlnGlnGlnGlnIleGlnGlnGlnAlaArgGluLeuLeuSerArg 1932  
OY 337 -----GAGCAAGTCTTATGACAGGAGTGAACAAGTTTTCGA 375  
DB 1933 GlnLysGlnAlaMetLysAlaGlnAlaGlyProValGlnGln-----GlnLeuLeuGln 1950  
OY 376 ATATCTGAGAGACCAATTTTGAAGAAACCAAGCCTTTAATCCAGTGTGACCAAGTGG 435  
DB 1951 GluThrGlnLysGlnMetLysGlnLysLeuGlnValGlnCysGlnAlaGlnValArg 1970  
OY 436 AAGGACTGAGAGCGAGCGAGCGACTTGAAGAAAGAACTTGCATCTCG-- 486  
DB 1971 AspAspLeuGlnLysGlnValLysAlaLeuGlnIleAspValGlnGlnValSerArg 1990  
OY 487 -----CAAGCAAAAGGGCCATTGAGAAACACATG----- 516  
DB 1991 PheIleGlnLeuGlnGlnGlnLysAsnThrGlnLeuMetArgPheArgGlnGlnAsnGln 2010  
OY 517 ---ATGAAAAAGGAATAACGAAGAAAGGAGTACATGGATCAAGATGTTG----- 567  
DB 2011 AlaLeuGlnLysGlnLeuGlnLysMetArgLysPheLeuAspGlnGlnAlaIleAspArg 2030  
OY 568 -----ATCTTGCTCAGAAATTTGCCCACTGAGAGGCCAGGTGGAAAG 612

DB 2031 GlnHisGlnLysArgSerValPheGlnGlnGlnIleGlnLysLeuGlnGlnLeuLysVal 2050  
OY 613 GTTACAAAGGAAAGATTTGACTATTAATCACTGAGGAAATTCAGACAGCTGGCT 672  
DB 2051 ValProArgPheGlnProIleSerGlnHisGlnThrArgGlnValGlu-----GlnLeuAla 2069  
OY 673 TCTCGGGAATGGATGTCACAAAGGTGTGAGAAATG-----CCCTATCAGCTG 723  
DB 2070 AsnHisLeuLysGlnLysThrAspLysCysSerGlnLeuLeuSerLysGlnGlnLeu 2089  
OY 724 AATTAACCAACATGGAAGAGATGAGGAGAAAGAGCAGACGAGCTTCAGAGCAAA 783  
DB 2090 GlnArgAspIleGlnGlnLysArgGlnLys-----GluIleGlnLysLeuGlnPheArgValArg 2108  
OY 783 ----- 783  
DB 2109 GluLeuGlnGlnAlaLeuLeuGlnLysArgLysHisPheGlnAlaValGlnAlaLysPro 2128  
OY 783 ----- 783  
DB 2129 GluLeuSerLeuGlnValGlnGlnAlaGlnLysArgSerAlaIleAspArgLysGluLys 2148  
OY 784 -----ACTAAC-----AGGATCTTGAATTAAGAT 810  
DB 2149 GluIleThrAsnLeuGlnGlnGlnGlnLeuGlnPheArgGlnGlnLeuGlnLysAsn 2168  
OY 811 CAGGAATATAGAAATTTGAGAAATAGACTGAT-----GAAACCAAAACACAC 858  
DB 2169 GluGlnValGlnGlnLeuGlnLysMetGlnLeuGlnIleGlnLysLysGlnSerThrThrArg 2188  
OY 859 TTGACACAGAGCAGCAGCAAGAGCCCTGGCCACAGACAGAGCTGCTGACACTAACGAA 918  
DB 2189 LeuGlnGlnLeuGlnGlnGlnLysLysLeuPheLysAspAspMetGluLysLeuGluLeu 2208  
OY 919 CTGCTGGCGCAATCTGAG-----CACCAACTGCACCTCACACAGATCTGAATA 966  
DB 2209 AlaIleLysGlnSerAspAlaMetSerThrGlnAspGlnHisValLeuPheGluLysPhe 2228  
OY 967 GCTCAACTCAGTCAGCAAAAGGATATCATATGATTAATTTGGA-----AAG 1014  
DB 2229 AlaGlnIleIleGlnGlnLysGlnValGluIleAspGlnLeuAsnGlnGlnValThrLys 2248  
OY 1015 TTACAGAGAGA-----AATGACAAATTTGAGCA----- 1044  
DB 2249 LeuGlnGlnGlnLeuLysIleThrThrAspAsnLysValIleGlnGlnLysAsnGlnLeu 2268  
OY 1045 -----CAGTGTGTCCAGCAT 1059  
DB 2269 IleArgAspLeuGlnLysThrGlnIleGlnLysLeuMetSerAspGlnGlnLysValLysArg 2288  
OY 1060 GGGAGAGTACATGAGAGATGAGCAAGCAAGGCTAAGCAGCTGGATGAAGCAGCCAGGCC 1119  
DB 2289 AsnArg-----GluGlnGlnIleIleGlnGlnLeuAsnGlnValIleGlnLys 2303  
OY 1120 ACAAGCCAGCAGCTGGTGCAGCTCTCAGCAGCAGCAGCAAGCTTCTCTGAGAGGCGAG 1179  
DB 2304 LeuGlnGlnGlnLeuLysAlaAsnIleGlnGlnLysThrSer-----MetAsnAlaHis 2320  
OY 1180 AGCCTGTGGAAGAGGTGAGCCGCTGCGGAGCCAGCTTA 1218  
DB 2321 SerLeuSerGlnGlnAlaAspSerLeuLysHisGlnLeu 2333

RESULT 8  
US-09-724-676A-68911  
Sequence 68911, Application US/09724676A  
GENERAL INFORMATION:  
APPLICANT: CompuGen LTD  
TITLE OF INVENTION: Variants of alternative splicing  
FILE REFERENCE: 129181.4 CompuGen  
CURRENT APPLICATION NUMBER: US/09/724,676A  
CURRENT FILING DATE: 2000-11-28  
NUMBER OF SEQ ID NOS: 97222  
SOFTWARE: PatentIn version 3.2

```
; SEQ ID NO 68911
; LENGTH: 3852
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-68911

Alignment Scores:
Prod. No.:      2, 18e-12      Length:      3852
Score:          268.50        Matches:      121
Percent Similarity: 41.23%    Conservative: 107
Best Local Similarity: 21.88% Mismatches:      162
Query Match:     9.83%        Indels:        163
DB:              5           Gaps:          22

US-09-502-945-1 (1-1552) x US-09-724-676A-68911 (1-3852)

QY      1 CTTCTGGATCATCCGAGAAAGCTTAATCTATGAGAAAGTGTAA--ATTGAG 57
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1797 MetLeuSnlleSerSerArgLeuGlnAlaValGluLysLeuGlnAlaIleSer 1816

QY      58 GAA-----TCCCAATTGAAGTTTGGAGAACGAC 87
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1817 GluThrSerSerGlnLeuGlnHisAlaLysValThrGlnThrGluLeuMetArgGluSer 1836

QY      88 TTAGCTGAATATATCAGAACTTGTGAAGATCTTAAGAGCACTAAAGCATTAAGAATT 147
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1837 PheArgGlnLysGlnGlnAlaThrGlnSerLeuLysGlnGlnGluLeuArgGluArg 1856

QY      148 CTTCTGGCTCTAATCTTAAACCGTGTGGTGTCTTTGTTGAATGCTCTACCAT 207
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1857 LeuHisGlnGlnSerArgAlaArgGlu-----GlnLeuAlaValGlnLeuSerLysAla 1874

QY      208 GAAGCTGTCTTCTCCCAACCCACTAATATGTCATATGCAGACCAACGAAAGCTGTT 267
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1875 GluGlyAlaIleAspGlyTrpAlaAspGluLysThrLeu-----PheGluArgGlnIle 1892

QY      268 AAGAAGAGATGACTGTGATGCTGCACATA-----GTTCCGTAAG 309
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1893 GlnGluLysThrAspIleIleAspArgLeuGlnGlnLeuLeuLysAlaSerAsnArg 1912

QY      310 AGCAGCTTGCACATCCGACCAAGA----- 336
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1913 LeuGlnGlnLeuGlnAlaGlnGlnGlnGlnIleGlnGlnGlnArgGlnLeuLeuSerArg 1932

QY      337 -----GAAGCAAGTCTTATGAGAGCAAGCAAGTTTGCA 375
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1933 GlnLysGlnAlaMetLysAlaGlnAlaGlyProValGlnGln-----GlnLeuGln 1950

QY      376 ATATCTGAGAGAGCCAAATTTGAAAAAACCAAGGCTTTAATCCAGTGTGACAGATTGAG 435
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1951 GluThrGlnLysLeuMetLysGlnLysLeuGlnValGlnLysGlnAlaGlnLysValArg 1970

QY      436 AAGAGCTGGAGAGCGCGGAGCCACTTGAAGAAAGAACTTGCACTGAC----- 486
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1971 AspAspLeuGlnLysGlnValLysAlaLeuGlnIleAspValGlnGlnGlnValSerArg 1990

QY      487 -----CAGAAGAAAAAGGCCCATGAGAAAGCAG----- 516
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1991 PheIleGlnLeuGlnGlnGlnLysAsnThrGlnLeuMetLysPheuArgGlnGlnAsnGln 2010

QY      517 ---ATGAAGAAAGAAATACGAAAGAAAGAGAGTACATGGAGTCAAGATGTTG----- 567
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2011 AlaLeuLysGlnLeuGlnLysMetArgLysPheLeuAsnGlnGlnAlaIleAspArg 2030

QY      568 -----ATCTGTCTCAGAAATTTGCCCACTGAGGCCCAAGGTGGAAG 612
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2031 GlnHisGlnLysArgAspValPheGlnGlnGlnGlnLysLeuGlnGlnGlnVal 2050

QY      613 GTTACAAGAAAGAAATTCATTAATCAACTGAGAAATTCAAACCGAGCTGGCT 672
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2051 ValProArgPheGlnProIleSerGlnHisGlnThrArgGlnValGlu-----GlnLeuAla 2069

QY      673 TCTCGGAATATGATGTACAAAGTGTGTGAGAAATG-----CGCTATCAGCTG 723
```

```
Db      2070 AsnHisLeuLysGlnLysThrAspLysCysSerGlnLeuLeuSerLysGlnGlnLeu 2089
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      724 AATTAACCAACATGAGAAAGATGAGCAAGAAAGAGCAGACAGATTCAAGCAAAA 783
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2090 GlnArgAspIleGlnGlnLysArgAsnGln-----GlnIleGlnLysLeuGlnPheArgValArg 2108

QY      783 ----- 783
Db      2109 GluLeuGlnGlnAlaLeuLeuGlnLysAspArgLysHisPheGlnAlaValGlnAlaLysPro 2128

QY      783 ----- 783
Db      2129 GluLeuSerLeuGlnValGlnLeuGlnAlaGlnLysArgAspAlaIleAspArgLysGlnLys 2148

QY      784 -----ACTAAC-----AGGATCTTGAATTAAGAT 810
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2149 GluThrAsnLeuGlnGlnGlnGlnLeuGlnGlnPheArgGlnGlnGlnLysAsn 2168

QY      811 CAGGAATATAGAAATATGAGAAATGAGACTGGAT-----GAAGCAAAACACAC 858
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2169 GluGlnValGlnGlnGlnHisMetGlnLeuGlnIleGlnLysLysGlnSerThrThrArg 2188

QY      859 TTGCACAGCAGCAGCAGCAGAGCCCTGCGCCAGAGAGAGTGCCTGAGACTAACAGAA 918
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2189 LeuGlnLysLeuGlnGlnGlnLysAsnLysLeuPheLysAspAspMetGlnLysLeuGlnLys 2208

QY      919 CTGCTGGCGCAATCTGAG-----CACCACTGCACCTCACACAGATCTGAATA 966
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2209 AlaIleLysGlnSerAspAlaMetSerThrGlnAspGlnHisValLeuPheGlnLysPhe 2228

QY      967 GCTCACTCAGTCAAGAAAAAGATATACATATGATTAATTTGGGA-----AAG 1014
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2229 AlaGlnIleIleGlnGlnLysGlnValGlnIleAspGlnLeuAsnGlnGlnValThrLys 2248

QY      1015 TTACAGAGAGA-----AATGAATAATTGAGAGAA----- 1044
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2249 LeuGlnGlnGlnLeuLysIleThrThrAspAsnLysValIleGlnGlnLysAsnGlnLeu 2268

QY      1045 -----CAGTGTCTCCAGCAT 1059
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2269 IleArgAspLeuGlnThrGlnIleGlnLysLeuMetSerAspGlnGlnLysValLysArg 2288

QY      1060 GGGAGATATACATGAGCAGATGAGCAAGCAAGCTAAGCAGCTGATAGCAGACAGCGCC 1119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2289 AsnArg-----GlnGlnGlnIleGlnGlnLeuAsnGlnValIleGlnLys 2303

QY      1120 ACAGCCAGCAGCTGTGTCAGCTCTTCAGCAGACAGACAGCACTTCTCTGAGAGGCGAG 1179
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2304 LeuGlnGlnGlnLeuAlaAsnIleGlnGlnLysThrSer-----MetAsnAlaHis 2320

QY      1180 AGCTGTGCGAAGAGTGTGACCGGCTCGGACCCAGTTA 1218
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2321 SerLeuSerGlnGlnAlaAspSerLeuLysHisGlnLeu 2333

RESULT 9
US-09-724-676A-68912
; Sequence 68912, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68912
; LENGTH: 3852
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-68912

Alignment Scores:
```

Pred. No.: 2.18e-12 Length: 3852  
 Score: 268.50 Matches: 121  
 Percent Similarity: 41.23% Conservative: 107  
 Best Local Similarity: 21.88% Mismatches: 162  
 Query Match: 9.83% Indels: 163  
 DB: 5 Gaps: 22

US-09-502-945-1 (1-1552) x US-09-724-676A-68912 (1-3852)

OY 1 CTTTGATGATGATCCGAGAGCAATAAATCTTATGAGAAAGTGTGA---ATTGAG 57  
 DB 1797 MetLeuAsnHisSerSerArgLeuGlnAlaValGluLysLeuGlnAlaHisSer 1816  
 OY 58 GAA-----TCCCAATTGAGATTGTTGAGAGACGAC 87  
 DB 1817 GluThrSerSerGlnLeuGlnHisAlaLysValThrGlnThrGluLeuMetArgLysSer 1836  
 OY 88 TTAAGCTGAATTCAGAGAACTTGAGATCTTAAAGACAACTAAAGCATTAAGCATTT 147  
 DB 1837 PheArgGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 1856  
 OY 148 CTTTGGCTGCTAATTAATCTGTAACCGTGTGGTCTTGTGTAATGTGCTGACAT 207  
 DB 1857 LeuHisGlnLysSerArgAlaArgGlu-----GlnLeuAlaValGluLeuSerLysAla 1874  
 OY 208 GAAGCTGTTCTTCCCAAAACCATACTAATGTTCAATGACAGCATCGAAGACTGTT 267  
 DB 1875 GlnGlyValIleAspGlyLysAlaSerGlnLysThrLeu-----PheGlnArgGlnIle 1892  
 OY 268 AAGAAAGAGATGACTGATGTCGACATA-----GTTCCCGTAAGC 309  
 DB 1893 GlnGlnLysThrAspIleLeuAspArgGlnGlnLysLeuLysCysAlaSerAsnArg 1912  
 OY 310 AGCAGCTTGGCAGATACGACCAAGA----- 336  
 DB 1913 LeuGlnGlnLysGlnAlaGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 1932  
 OY 337 -----GAAGCAAGTCTTATGAACAGAGTGAACAACTTTTGCAA 375  
 DB 1933 GlnLysGlnAlaMetLysAlaGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 1950  
 OY 376 ATATCTGAGGAAGCAATTTTGAAGAAACCAAGGCTTTAATCCAGTGTGACAGTTGAGC 435  
 DB 1951 GlnThrGlnLysLeuMetLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 1970  
 OY 436 AAGGAGCTGAGAGCGAGCGAGCACTTGAAAGAAAGCACTGCAATCTGAG----- 486  
 DB 1971 AspAspLeuGlnLysGlnLysAlaLeuGlnLysAlaLeuGlnLysValSerArg 1990  
 OY 487 -----CAAGAGAAAGGGCCATTGAGAAAGACATG----- 516  
 DB 1991 PheIleGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 2010  
 OY 517 ---ATGAAAGAAAGAAATTAAGAAAGAGGAGTACATGGAATCAAGATGTG----- 567  
 DB 2011 AlaLeuGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 2030  
 OY 568 -----ATCTTGTCTCAGATATTTGCCCACTGAGAGCCAGCGTGGAAAG 612  
 DB 2031 GlnHisGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 2050  
 OY 613 GTTACAAAGGAAAGATTTACCTATTAATCACTGAGGAAATTCAAAGCCAGCTGGCT 672  
 DB 2051 ValProArgPheGlnProIleSerGlnHisGlnThrArgGlnValGlu---GlnLeuAla 2069  
 OY 673 TCTCGGGAATGATGATCACAAGGTGTGAGAAATG-----CGCTATACGCTG 723  
 DB 2070 AsnHisLeuLysGlnLysThrAspLysCysSerGlnLeuLeuSerLysGlnLysGlnLys 2089  
 OY 724 AATAAACAACATGAG 783  
 DB 2090 GlnArgAspIleGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 2108

OY 783 ----- 783  
 DB 2109 GluLeuGlnLysAlaLeuLeuGlnLysAspArgLysHisPheGlnValGlnLysPro 2128  
 OY 783 ----- 783  
 DB 2129 GluLeuSerLeuGlnValGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 2148  
 OY 784 -----ACTAAC-----AGGAGCTTGAATTAAGAT 810  
 DB 2149 GluIleThrAspLeuGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 2168  
 OY 811 CAGGAATTAAGAAATTTGGAATTAAGATGAT-----GAAAGCAACACACAC 858  
 DB 2169 GlnGlnValGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 2188  
 OY 859 TTGGAACAGGACACAG 918  
 DB 2189 LeuGlnGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 2208  
 OY 919 CTGCTGGCGCAATCTGAG-----CACCAACTGCACCTGACACAGATCTGAAATA 966  
 DB 2209 AlaIleLysGlnLysSerAspAlaMetSerThrGlnAspGlnHisValLeuPheGlnLysPhe 2228  
 OY 967 GCTCAACTGATGCAAGAAAAAGATATACATATGATTAATTTGGA-----AAG 1014  
 DB 2229 AlaGlnIleIleGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 2248  
 OY 1015 TTACAGACAGAGA-----AATGAAGATTTGAGAGA----- 1044  
 DB 2249 LeuGlnGlnLysGlnLysLysLeuThrThrAspAsnLysValIleGlnLysGlnLysGlnLys 2268  
 OY 1045 -----CAGTGTCCAGCAT 1059  
 DB 2269 IleArgAspLeuGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 2288  
 OY 1060 GGGAGAT 1119  
 DB 2289 AsnArg-----GlnGlnGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 2303  
 OY 1120 ACAGCCGAGCAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1179  
 DB 2304 LeuGlnGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 2320  
 OY 1180 AGCCTGTGGAAGAGTGGAGCCGCTGCGAGCCAGTTA 1218  
 DB 2321 SerLeuSerGlnLysAlaAspSerLeuLysHisGlnLeu 2333  
 RESULT 10  
 US-09-724-676A-68913 ; Sequence 68913, Application US/09724676A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Compugen LTD  
 ; TITLE OF INVENTION: Variants of alternative splicing  
 ; FILE REFERENCE: 129181.4 Compugen  
 ; CURRENT APPLICATION NUMBER: US/09/724.676A  
 ; NUMBER OF SEQ ID NOS: 2000-11-28  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 68913  
 ; LENGTH: 3852  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-724-676A-68913  
 Alignment Scores:  
 Pred. No.: 2.18e-12 Length: 3852  
 Score: 268.50 Matches: 121  
 Percent Similarity: 41.23% Conservative: 107  
 Best Local Similarity: 21.88% Mismatches: 162  
 Query Match: 9.83% Indels: 163  
 DB: 5 Gaps: 22

US-09-502-945-1 (1-1552) x US-09-724-676A-68913 (1-3852)

QY 1 CTTCTGATGATCCGAGAACCTAAACTTACTATATGAGAAAGTGTGAA--ATTGAG 57  
 Db 1797 MetLeuasnIleSerSerArgLeuGlnAlaValGluLysLeuLeuGlnAlaIleSer 1816

QY 58 GAA-----TCCCAATTGAAGTTTGTGAGAACGAC 87  
 Db 1817 GluThrSerSerGlnLeuGlnHisAlaLysValThrGlnThrGlnLeuMetArgGluSer 1836

QY 88 TTACCTGATATTCAGAGAACTTGTGAAGATCTTAAAGCAACTAAAGCAATTAAGATTT 147  
 Db 1837 PheArgGlnLysGlnGlnAlaThrGluSerLeuLysCysGlnGlnGluLeuArgGluArg 1856

QY 148 CTTCTGGCTGCTTAATTAATCTTAACCGTGTCTTGTGAAATGTCACAGCAT 207  
 Db 1857 LeuHisGluGluSerArgAlaArgGlu-----GlnLeuAlaValGluLeuSerLysAla 1874

QY 208 GAAGCTGTTCTTCCCAAAACCCATTAATGTTCATATGACAGACATCGAAAGACTGGTT 267  
 Db 1875 GluGluValIleAspGlyTyrAlaAspGluLysThrLeu-----PheGluArgGlnIle 1892

QY 268 AAAAGAAAGAGATGACTTGATGTCTGCACATA-----GTTCCGCTAAGG 309  
 Db 1893 GlnGluLysThrAspIleIleAspArgLeuGlnGluLeuLysCysAlaSerAsnArg 1912

QY 310 AGCAGCTTGGCAGATACCGCACAAGA----- 336  
 Db 1913 LeuGlnGluLeuGlnAlaGlnGlnGlnGlnIleGlnGlnGluArgGluLeuLeuSerArg 1932

QY 337 -----GAAGCAAGTCTTATGACAGCTGAAACAGTTTGTGCA 375  
 Db 1933 GlnLysGluAlaMetLysAlaGlnAlaGlyProValGlnGln-----GlnLeuLeuGln 1950

QY 376 ATATCTGAGAGACCAATTTTGAAGAAAACCAAGCCTTTATTCAGCTGTGACCAGTGTGAG 435  
 Db 1951 GluThrGluLysLeuMetLysGlnLysLeuGlnValGlnCysGlnAlaGlnLysValArg 1970

QY 436 AAGAGCTGAGAGAGCAGCGGAGCTGTGAAAGAAAGAACTTGATCTGAC----- 486  
 Db 1971 AspAspLeuGlnLysGlnLeuValLysAlaLeuGlnIleAspValGlnGlnGlnValSerArg 1990

QY 487 -----CAAGAGAAAAGGCCATGTGAGAAAGCATG----- 516  
 Db 1991 PheIleGluLeuGlnGlnGlnGlnLysAsnThrGluLeuMetAspLeuArgGlnGlnAsnGln 2010

QY 517 ---ATGAAAAGAAATTAACAAAGAAAGGAGTACATGGATCAAGATTTG----- 567  
 Db 2011 AlaLeuGlnLysGlnLeuGlnLysMetArgLysPheLeuAspGlnGlnAlaIleAspArg 2030

QY 568 -----ATCTGTCTCAGAAATATGCCCAACTGAGAGGCCAGGTGGAAGAAG 612  
 Db 2031 GlnHisGluArgAspValPheGlnGlnGlnIleGlnLysLeuGlnGlnGlnLysVal 2050

QY 613 GTTACAAAGAAAGATTTTCACTATTATCACTGAGAGAAATTCAAAGCCAGCTGGCT 672  
 Db 2051 ValProArgPheGlnProIleSerGlnHisGlnThrArgGluValArg---GlnLeuAla 2069

QY 673 TCTCGGGAATGATGTCACAAAGGTGTGTGGAATATG-----CGTTCATCAGCTG 723  
 Db 2070 AsnHisLeuLysGlnLysThrAspLysCysSerGlnLeuLeuSerLysGlnGlnLeu 2089

QY 724 AATTAACCAACATGAGAGATGAGCAGAAAAGGAGCAGAGAGCTTACAGCAAAA 783  
 Db 2090 GlnArgAspIleGlnGlnArgAsnGln---GluIleGlnLysLeuGlnPheArgValArg 2108

QY 783 ----- 783  
 Db 2109 GluLeuGlnGlnAlaLeuLeuGlnAspArgLysHisPheGlnAlaValGlnAlaLysPro 2128

QY 783 ----- 783  
 Db 2129 GluLeuSerLeuGlnValGlnLeuGlnAlaGlnLysArgAspAlaIleAspArgLysGlnLys 2148

QY 784 -----ACTAAC-----AGGAGCTTGAATTAAGAT 810  
 Db 2149 GluIleThrAsnLeuGlnGlnGlnLeuGlnGlnPheArgGlnGlnGlnLysAsn 2168

QY 811 CAGGAATATGAGAAATTTGAGATGAACTGAT-----GAAGCAACCAACAC 858  
 Db 2169 GluGluValGlnGlnGlnLeuHisMetGlnLeuGlnIleGlnLysGlnSerThrArg 2188

QY 859 TTGAAACAGAGCAAGCAAGAGCAAGCCCTGGCCAGAGAGAGTGTGAGCTAAACAGA 918  
 Db 2189 LeuGlnGlnLysGlnGlnGlnAlaAsnLysLeuPheLysAspAspMetGlnLysLeuGluLeu 2208

QY 919 CTTCTGGGCAATCTGAG-----CACCAACTGCACTCACACAGATCTGAATA 966  
 Db 2209 AlaIleLysGluSerAspAlaMetSerThrGlnAspGlnHisValLeuPheGlyLysPhe 2228

QY 967 GCTCAACTCACTCAAGAAAAGGTATACATATGATTAATTGGA-----AAG 1014  
 Db 2229 AlaGlnIleIleGlnGlnLysGlnValGlnLysGlnLeuAsnGlnGlnValThrLys 2248

QY 1015 TTACAGACAGA-----ATGAAAGATTTGAGAGAA----- 1044  
 Db 2249 LeuGlnGlnGlnLysLysIleThrThrAspAsnLysValIleGlnGlnLysAsnGlnLeu 2268

QY 1045 -----CAGTGTCCAGCAT 1059  
 Db 2269 IleArgAspLeuGluThrGlnIleGlnCysLeuMetSerArgGlnGlnCysValLysArg 2288

QY 1060 GGGAGAGTACATGAGACGATGAGCAAGGCTAAGCAGCTGATTAAGCAGACAGGCC 1119  
 Db 2289 AsnArg-----GlnGlnGlnIleGlnGlnGlnLeuAsnGlnValIleGlnLys 2303

QY 1120 ACAGCCCAAGCAGCTGTGTCACCTCTCAGCAAGCAAGCAAGCTTCTCTGGAAGGCA 1179  
 Db 2304 LeuGlnGlnGlnLysAlaAsnIleGlyGlnLysThr-----MetAsnAlaHis 2320

QY 1180 AGCCTGTGCAAGAGTGTGACCGGCTGCGAGCCAGCTTA 1218  
 Db 2321 SerLeuSerGlnGlnAlaAspSerLeuLysHisGlnLeu 2333

RESULT 11  
 US-09-724-676-68923  
 ; Sequence 68923, Application US/09724676  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Compugen LTD  
 ; TITLE OF INVENTION: Variants of alternative splicing  
 ; FILE REFERENCE: 129181.4 Compugen  
 ; CURRENT APPLICATION NUMBER: US/09/724,676  
 ; NUMBER OF SEQ. ID NOS: 2000-11-28  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO: 68923  
 ; LENGTH: 3864  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-724-676-68923

Alignment Scores:  
 Pred. No.: 2,18e-12 Length: 3864  
 Score: 268.50 Matches: 121  
 Percent Similarity: 41.23% Conservative: 107  
 Best Local Similarity: 21.88% Mismatches: 162  
 Query Match: 9.83% Indels: 163  
 DB: Gaps: 22

US-09-502-945-1 (1-1552) x US-09-724-676-68923 (1-3864)

QY 1 CTTCTGATGATCCGAGAACCTAAACTTACTATATGAGAAAGTGTGAA--ATTGAG 57  
 Db 1809 MetLeuasnIleSerSerArgLeuGlnAlaValGluLysLeuLeuGlnAlaIleSer 1828

QY 58 GAA-----TCCCAATTGAAGTTTGTGAGAACGAC 87

```

Db 1829 GIUthSerSerGlnLeuGlnHisAlaLysValThrGlnThrGlnLeuMetArgLysUser 1848
OY 88 TTAGCTGAATATCAGAGAACTTGTGAAGATCTTAAAGACGATAAAGCAATT 147
Db 1849 PheArgGlnLysLeuGlnAlaThrGlnSerLeuLysCysGlnGlnLeuArgGlnLys 1868
OY 148 CTTCTGCTGCTATATCTGTAAACCGTGTGTGCTTTGTTGAATGTGCTCAGAT 207
Db 1869 LeuHisGlnLysSerArgAlaArgGln-----GlnLeuAlaValGlnLeuSerLysAla 1886
OY 208 GAAGCTGTCTTCCCAAAACCCATTAATGTTCAATGACAGACATCGAAAGATGCTT 267
Db 1887 GlnGlnValIleAspGlnLysAlaSerGlnLysThrLeu-----PheGlnArgGlnIle 1904
OY 268 AAAAGAAAGATGACCTTGAATGTCGACAT-----GTTTCCGTAAG 309
Db 1905 GlnGlnLysThrAspIleLeuAspArgLeuGlnGlnLeuLysCysAlaSerAsnArg 1924
OY 310 AGCAGCTTGACATACGACGACAAAGA----- 336
Db 1925 LeuGlnGlnLysLeuAlaGlnGlnGlnIleGlnGlnLysArgLysLeuLeuSerArg 1944
OY 337 -----GAGCAAGTGCCTTATGACAGGTGAACAGATTGCGAA 375
Db 1945 GlnLysGlnAlaMetLysAlaGlnAlaGlnLysProValGlnLys-----GlnLeuLeuGln 1962
OY 376 ATATCTGAGAAAGCAATTTGAAAAAACCAAGGCTTTAATCCAGTGCAGTGCAGG 435
Db 1963 GlnThrGlnLysLeuMetLysGlnLysLeuGlnValGlnCysGlnAlaGlnLysValArg 1982
OY 436 AAGGACCTGAGAGGACGCGGAGGACCTTGAAGAAAGACTTCATCTCAG----- 486
Db 1983 AspAspLeuGlnLysGlnValLysAlaLeuGlnIleAspValGlnLysGlnValSerArg 2002
OY 487 -----CAAGGAAAGAGGCGCATGAGAAAGACATG----- 516
Db 2003 PheIleGlnLysGlnGlnLysAsnThrGlnLeuMetAspLeuArgGlnGlnAsnGln 2022
OY 517 ---ATGAAAGAAAGAAATTAACGAAGAAGGAGATACATGAGATCAAAAGATGTTG----- 567
Db 2023 AlaLeuGlnLysGlnLeuGlnLysMetArgLysPheLeuAspGlnGlnAlaIleAspArg 2042
OY 568 -----ATCTGTCTCAGAATATGCCCCAACCTGAGGCCACGTGGAAAG 612
Db 2043 GlnHisGlnLysArgValPheGlnGlnGlnIleGlnLysGlnGlnLysVal 2062
OY 613 GTTACAAAGAAAGATTTACGATTAATCACTGAGAAATTCAAAGCCAGCTGGCT 672
Db 2063 ValProArgPheGlnProIleSerGlnHisGlnThrArgGlnValGln--GlnLeuAla 2081
OY 673 TCTCGGGAATGATGTCAAAAGTGTGTGAGAAATG-----GCTTATCAGCTG 723
Db 2082 AsnHisLeuLysGlnLysThrAspLysCysSerGlnLeuLeuSerLysGlnGlnLeu 2101
OY 724 AATAAACAACACATGAGAGATGAGGCAAAAGAGACACAGAGATTCAGAGCAAAA 783
Db 2102 GlnArgAspIleGlnLysArgGlnLysGlnLysGlnLysLeuPheArgValArg 2120
OY 783 ----- 783
Db 2121 GlnLeuGlnGlnAlaLeuLeuGlnLysAspArgLysHisPheGlnValAlaGlnLysPro 2140
OY 783 ----- 783
Db 2141 GlnLeuSerLeuGlnValGlnLeuGlnAlaGlnLysArgAspAlaIleAspArgLysGlnLys 2160
OY 784 -----ACTAAC-----AGGATCTTGAATTAAGAT 810
Db 2161 GlnIleThrGlnLeuGlnLysGlnLysGlnLysPheArgGlnLysGlnLysLysAsn 2180
OY 811 CAGGAATGAGAAATGAGAACTGAGAT-----GAAAGCAAAACAAC 858

```

```

Db 2181 GlnGlnValGlnGlnLeuHisMetGlnLeuGlnIleGlnLysGlnSerThrThrArg 2200
OY 859 TTGACAGAGAGCAGCAGAGAGCAGCCCTGCGCAGAGAGATGCTGAGACTTAACAGAA 918
Db 2201 LeuGlnGlnLysGlnGlnLysAsnLysLeuPheLysAspMetGlnLysLeuGlnLys 2220
OY 919 CTGCTGGGGAATCTGAG-----CACCAACTGCACCTCACCAATCTGAATA 966
Db 2221 AlaIleLysGlnSerAspAlaMetSerThrGlnAspGlnHisValLeuPheGlnLysPhe 2240
OY 967 GCTCAACTCAGTCAACAAAAAGATATACATATGATTAATTTGGA-----AAG 1014
Db 2241 AlaGlnIleIleGlnGlnLysGlnValGlnIleAspGlnLeuAsnGlnGlnValThrLys 2260
OY 1015 TTACAGAGAA-----AATGAGAATTTGAGGAA----- 1044
Db 2261 LeuGlnGlnGlnLysLysIleThrThrAspAsnLysValIleGlnLysAsnGlnLeu 2280
OY 1045 -----CAGTGTGTCAGCAT 1059
Db 2281 IleArgAspLeuGlnThrGlnIleGlnLysLeuMetSerAspGlnGlnLysValLysArg 2300
OY 1060 GGGAGAGTACATGACAGATGAGCAAGGCTTAAGCAGCTGATTAAGCACAGCCAGGCC 1119
Db 2301 AsnArg-----GlnGlnGlnIleGlnGlnLysGlnGlnValIleGlnLys 2315
OY 1120 ACAGCCAGCAGCTGCTGACGCTCTCCAGCAACAGACACAGCTTCTCTGAGAGCGAG 1179
Db 2316 LeuGlnGlnGlnLeuAlaAsnIleGlnGlnLysThrSer-----MetAsnAlaHis 2332
OY 1180 AGCCTGTGCGAGAGGTGACCGCGCTCGGACCCAGTTA 1218
Db 2333 SerLeuSerGlnLysAlaAspSerLeuLysHisGlnLeu 2345

```

RESULT 12

```

US-09-724-676-68924
; Sequence 68924, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724, 676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68924
; LENGTH: 3864
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-724-676-68924

```

Alignment Scores:

Pred. No.:	2,18e-12	Length:	3864
Score:	268.50	Matches:	121
Percent Similarity:	41.23%	Conservative:	107
Best Local Similarity:	21.86%	Mismatches:	162
Query Match:	9.83%	Indels:	163
DB:	5	Gaps:	22

```

US-09-502-945-1 (1-1552) x US-09-724-676-68924 (1-3864)
OY 1 CTTCTGATGATCGAGAGAACTTAAGATTTATGAGAAAAAGTGTGAA---ATTGAG 57
Db 1809 MetLeuAsnIleSerSerArgLeuGlnAlaAlaValGlnLysLeuGlnAlaIleSer 1828
OY 58 GAA-----TCCCAATTTGAGTTTGTGAGAGACG 87
Db 1829 GlnThrSerSerGlnLeuGlnHisAlaLysValThrGlnThrGlnLeuMetArgLysUser 1848
OY 88 TTAGCTGAATATCAGAGAACTTGTGAAGATCTTAAAGACGATAAAGCAATT 147
Db 1849 PheArgGlnLysGlnGlnAlaThrGlnSerLeuLysCysGlnGlnLysLeuArgGlnLys 1868

```

[illegible]

```

Db 2221 A|a|l|e|y|s|g|u|s|e|r|a|s|p|a|l|a|e|t|S|e|r|T|h|C|l|a|s|p|c|l|n|h|i|s|v|a|l|e|u|r|P|h|e|c|t|y|L|y|s|P|h|e 2240
QY 967 G|C|T|C|A|A|C|A|C|A|C|T|C|A|A|A|A|A|A|A|G|G|T|A|T|A|C|A|T|A|T|G|A|T|A|A|T|T|G|G|A|-----A|A|G 1014
Db 2241 A|a|e|g|i|n|i|e|i|l|e|g|i|n|g|i|u|y|s|g|i|v|a|l|c|i|l|i|e|a|s|p|i|c|i|n|e|u|a|n|s|i|n|c|i|u|b|i|n|v|a|l|h|r|y|s 2260
QY 1015 T|T|A|C|A|G|A|C|A|G|A|-----A|A|T|G|A|A|G|A|T|T|G|G|A|G|A|-----1044
Db 2261 l|e|u|g|i|n|g|i|n|t|e|u|l|y|s|i|e|t|h|r|t|h|r|a|s|p|a|n|b|i|l|y|v|a|i|l|e|g|i|u|c|i|u|y|s|a|n|g|i|u|e 2280
QY 1045 -----C|A|G|T|G|T|C|C|A|G|C|A|T 1055
Db 2281 l|l|e|a|r|g|a|s|p|i|e|u|g|i|u|t|h|r|i|n|g|i|l|l|e|g|i|c|y|s|l|e|u|e|t|S|e|r|a|s|p|i|c|i|n|c|i|u|b|i|y|s|v|a|l|y|a|r|g 2300
QY 1060 G|G|G|A|G|A|G|T|A|C|A|G|A|C|A|T|G|A|A|G|C|A|A|A|G|G|C|T|A|A|G|C|A|C|G|T|G|A|T|A|A|G|C|A|C|G|C 1119
Db 2301 A|s|h|a|r|g|-----G|i|u|c|i|u|g|i|i|l|l|e|g|i|u|b|i|e|u|a|n|s|i|n|c|i|u|b|i|l|l|e|g|i|u|y|s 2315
QY 1120 A|C|A|G|C|C|A|G|C|A|G|T|G|T|G|C|A|G|C|T|C|T|C|A|G|A|C|A|C|A|G|C|A|C|G|T|C|T|C|T|G|A|G|A|G|C|A|G 1179
Db 2316 l|e|u|g|i|n|g|i|n|t|e|u|a|i|a|s|i|l|l|e|g|i|l|y|g|i|l|y|t|h|r|S|e|r|-----M|e|t|a|s|i|n|a|i|s 2333
QY 1180 A|G|C|G|T|G|C|G|A|A|G|G|T|G|G|A|G|C|G|G|C|T|G|C|G|G|A|C|C|A|C|A|T|A 1218
Db 2333 S|e|r|l|e|u|S|e|r|G|i|u|a|i|a|a|s|p|S|e|r|l|e|u|Y|h|i|s|i|n|d|e|u 2345

RESULT 13
US-09-724-676-68925
; Sequence 68925, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68925
; LENGTH: 3864
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-68925

Alignment Scores:
Pred. No.: 2,18e-12 Length: 3864
Score: 268.50 Matches: 121
Percent Similarity: 41.23% Conservative: 107
Best Local Similarity: 21.88% Mismatches: 162
Query Match: 9.83% Indels: 163
DB: 5 Gaps: 22

US-09-502-945-1 (1-1552) x US-09-724-676-68925 (1-3864)
QY 1 C|T|T|G|G|A|T|G|C|A|T|C|C|G|A|A|G|A|C|T|T|A|A|C|T|T|A|T|G|A|G|A|A|A|G|T|G|A|---A|T|G|A|G 57
Db 1809 M|e|t|l|e|u|a|s|i|l|l|e|S|e|r|S|e|r|a|r|g|l|e|u|G|i|n|a|i|a|i|a|l|a|l|e|u|b|i|y|s|l|e|u|e|u|G|i|u|a|i|l|e|S|e|r 1828
QY 58 G|A|A|-----T|C|C|A|A|T|T|G|A|A|G|T|T|T|T|G|G|A|A|C|A|C 87
Db 1829 G|u|t|h|r|S|e|r|S|i|n|l|e|u|G|i|n|i|s|i|a|i|a|l|y|s|v|a|i|t|h|r|G|i|n|h|i|c|i|l|e|u|e|u|e|r|S|i|n|c|i|u|b|i|y|S|e|r 1848
QY 88 T|T|A|G|C|A|A|T|T|C|A|G|A|A|C|A|C|T|T|G|A|G|A|G|A|C|T|T|A|A|G|A|C|A|C|A|T|A|A|G|C|A|T|A|A|G|A|A|T|T 147
Db 1849 P|h|e|a|r|g|i|n|y|s|g|i|n|g|i|u|a|i|l|h|i|n|g|i|u|s|e|r|l|e|u|b|i|y|c|y|s|g|i|n|c|i|u|b|i|e|u|a|r|g|i|u|r|g 1866
QY 148 C|T|T|G|C|G|C|T|A|T|A|C|T|T|G|T|A|A|C|C|G|T|G|T|G|G|T|T|G|T|T|G|A|A|T|G|T|C|T|C|A|G|A|T 207
Db 1869 l|e|u|h|i|s|i|n|g|i|u|s|e|r|a|r|g|i|a|i|a|r|g|i|u|-----G|i|n|l|e|u|i|a|i|a|i|l|e|u|b|i|e|r|y|s|i|a|l 1886
QY 208 G|A|A|C|G|T|G|T|T|C|C|A|A|C|C|A|C|A|T|A|A|G|T|T|C|A|T|A|T|G|C|A|A|C|A|C|A|C|A|C|T|G|A|A|G|A|C|T|G|T 267
Db 1887 G|i|u|g|i|y|a|i|i|l|e|a|s|p|i|g|i|y|r|i|a|i|a|a|s|p|i|c|i|u|b|i|y|S|h|l|e|u|-----P|h|e|i|u|r|g|i|n|l|i|e 1904

```

OY	268	AAGAAGACAGATGACTTGATGTGCACCTA-----GTTCCCGTAAGG	309
Dd	1905	GinGlutylsThrAspIleLeaspArgLueGlnGluLeucCysAlasertasnrg	1922
OY	310	AGCACCTTGGCCGATACGCCACAAGA-----	336
Dd	1925	LengIngluleuglUValaglInglnInglleGlnclUglUargLgLUleuLeuseratrg	1944
OY	337	-----GAAGCAAGCGCTTAGACAGCGTGAAACAAGCTTTTGCA	375
Dd	1945	GlnLysglUAlametylSalagUAlaGlyProvalGlnUdn-----GlnLeuEngLn	1967
OY	376	ATAATCTGAGAAAGCCAAATTTGAAAAAACCAAGCTTTATCCAGTCGACCAGCTTGAG	435
Dd	1963	GlnThrGlnLysLeumetylSglnLysLueGlnUValGlcncsglnlaGlnLysValnrg	1982
OY	436	AAGAGCGTGGAGAGCGACGCCGAGCCACTTGAAAAGAAAGAACTTGCAATCTAG-----	486
Dd	1983	AspAspLeuglnLysGlnUValLysAlaLeuGlnLileaspArgLcUgnUValLysertArg	2002
OY	487	-----CAAGCAAAAGAGGCCCATTTGACAAAGACATG-----	516
Dd	2003	PheIleGlnLueGlnUgnLglnLglnLysAsnthrGlnLeuMetCapLeunrgLnlInsngln	2022
OY	517	---ATGAAAAAGCAATPACGAAGAAGAAGGCACTACATGGAGTACAAAGATGTTG-----	567
Dd	2023	AlaleuglnLysGlnLeuglnLysMetArgLysPheLeuAspLcUdnAlaLileaspArg	2042
OY	568	-----ATCTTGTCACAAATATTGCCCACTGGAGGCCCGCAGGTGGAAAG	612
Dd	2043	GlnUnlsGlnUargAspValPheGlnGlnGlnUlleGlnLysLueGlnUgnLinsUVal	2066
OY	613	GTTACAAAGGAAAGATTTTCACTATTATCAACTGGAGAAATTTCAAAGCCAGCTGGCT	672
Dd	2063	ValProArgPheGlnProIleSerGlnUnlsGlnThrArgGlnUValGlnU----GlnLeuAla	2081
OY	673	TCTCGGAAATGAGATGCACAAGAGCTGTGGGAATG-----CGCATACAGCTG	723
Dd	2082	AsnhlsleuLysGlnLysThrAspLysCysSerGlnLeuLeuLeuserLysGlnUlnLeu	2101
OY	724	AATNAAMCCAACATGACGAAGGATGAGGCGAAGAAAGGACAGCAGAGTCTGACAGCAAA	783
Dd	2102	GlnArgAspIleGlnGlnUArgasnGlnU--GlnUleGlnLysLcUdnPheArgUalArg	2126
OY	783	-----	783
Dd	2121	GlnLeuglnUlnAlaLeuLeuGlnUAspArgLysHslPheGlnAlaValGlnUAlalsPro	2140
OY	783	-----	783
Dd	2141	GlnLeuSerLeuGlnUValGlnLeuGlnAlaGlnUArgAspAlaIleaspArgLysGlnLys	2160
OY	784	-----ACTAAC-----AGGAGTCTGAATTTAAGAT	810
Dd	2161	GlnUleThrAsnLeuGlnUgnLglnLeuGlnUlnPheArgLcUgnUdnLeuLnsLysasn	2180
OY	811	CAGGAATACGAAATTGAGATGATGAAGCTGGAT-----GAAAGCAACAACAAC	858
Dd	2181	GlnUlnUValGlnGlnLnsUlsMeGlnLeuGlnUlleGlnLysLysGlnUserThrTrArg	2200
OY	859	TTTGACACGAGAGCGACGAAGCGACGCCCTGGCCAGAGACAGAGCTGCGACTAACAADA	918
Dd	2201	LenglnGlnLueGlnUgnLglnUlnsnLysLuePheLysAspAspMecUlnLysLcUlnLyeu	2220
OY	919	CTGCTGGGCGAATCTGAG-----CACCAACTGCACCTCACCGATCTGAATA	966
Dd	2221	AlaIleLysGlnUserAspAlaMetSetrThGlnAspGlnHslValLeuPheGlnLysPhe	2240
OY	967	GCTCAACCTCAACGAAAAAGGTATACATATGATTAATTTGGGA-----AAG	1014
Dd	2241	AlaGlnIleIleGlnGlnLysGlnUValGlnUlleAspGlnLeuAsnGlnUlnValnLhrLys	2260

QY	1015	TTACAGAGGAA	-----	AATGAAGAA	TTGGAGAA	-----	1044			
Db	2261	LeuGInGInGInLeuLysIleThrThrAspAsnLysValIleGluLysAsnGluLeu					2280			
QY	1045	-----	-----	-----	-----	-----	1059			
Db	2281	IleArgAspLeuGluThrGlnIleGluLysLeuMetSerAspGlnGluLysValLysArg					2300			
QY	1060	GGGAGAGACTGTAGACAGATGAAGCAAGGCTTAAGCAGCTGGATGATAGCAAGCCAGGC					1119			
Db	2301	AsnArg					2315			
QY	1120	ACAGCCAGCAGCTGTGTGCAGCTCTCCAGCAACAACAGCTTCTCCGTGAGAGAGAG					1179			
Db	2316	LeuGInGInGluLeuAlaAsnIleGluGlnLysThrSer					2333			
QY	1180	AGCCTGTGGAGAGGTGACCCGGCTGCGGACCCAGTTA					1218			
Db	2333	SerLeuSerGlnGluAlaAspSerLeuLysIleGlnLeu					2345			
RESULT 14										
US-09-724-676A-68923	Sequence 68923, Application US/09724676A									
;	GENERAL INFORMATION:									
;	APPLICANT: Compugen LTD									
;	TITLE OF INVENTION: Variants of alternative splicing									
;	FILE REFERENCE: 129181.4 Compugen									
;	CURRENT APPLICATION NUMBER: US/09/724,676A									
;	CURRENT FILING DATE: 2000-11-28									
;	NUMBER OF SEQ ID NOS: 97222									
;	SOFTWARE: PatentIn version 3.2									
;	SEQ ID NO 68923									
;	LENGTH: 3864									
;	TYPE: PRF									
;	ORGANISM: Homo sapiens									
US-09-724-676A-68923										
Alignment Scores:										
Pred. No.:	2,18e-12	Length:	3864							
Score:	268.50	Matches:	121							
Percent Similarity:	41.23%	Conservative:	107							
Best Local Similarity:	21.88%	Mismatches:	162							
Query Match:	9.83%	Indels:	163							
DB:	5	Gaps:	22							
US-09-502-945-1 (1-1552) x US-09-724-676A-68923 (1-3864)										
QY	1	CTTGTGATGCATCCGAGAGCTAAACCTACTTATGAGAAAGTGTGA	---ATGAG	57						
Db	1809	MetLeuAsnIleSerSerArgLeuGlnAlaAlaValGlnLysLeuGlnAlaIleSer	1828							
QY	58	GAA	-----TCCCAATTAAGTTTTTGAAGACAC	87						
Db	1829	GlnThrSerSerGlnLeuGlnIleAlaLysValThrGlnThrGluLeuMetArgIleSer	1848							
QY	88	TTACGTGAATATCAGAGACACTTGTGAAGATCTTAAAGACCAACTAAAGCAATTAAGACTTT	147							
Db	1849	PheArgGlnLysGlnGlnAlaThrGlnSerLeuLysCysGlnGlnGlnLeuArgLysArg	1868							
QY	148	CTTGTGCGCTGAATACTTCTTAACCGGTGTGGTGGCTTTGTTGAATGTGCTACGAT	207							
Db	1869	LeuHisGlnGlnSerArgAlaArgGlu	1886							
QY	208	GAACCTGTCTTCTTCCCAACCCACTACTAATGTTCATATGCAACACATGGAAGACTGGTT	267							
Db	1887	GlnIleValIleAspGlnLysAlaAspGlnLysThrLeu	1904							
QY	268	AAAGAAAGAGATGACTGATGTGTGCACTA	-----GTTCCGTAAG	309						
Db	1905	GlnIleLysThrAspIleIleAspArgLeuGlnGlnLeuLysAlaSerAsnArg	1924							
QY	310	AGCAGCTTGCGACAGTACGACGAAAGA	-----	336						





Db 1963 glnthglnlyseumethylsglnlysglnvalglnlncysglnlagauglylvalarg 1982  
QY 436 AAGGACCTGAGAGCGAGCGAGCTTGAAAGAACTTGACATCTGAG----- 486  
Db 1983 ASPASPLEGlnlysglnvalglnlalaLeuGlnlIleaspralgluglnvalIleSera 2002  
QY 487 -----CAAGAGAAAGGGCCATTGAGAAAGACATG----- 516  
Db 2003 PheIleGlnleuGlnlGlnlulysasnlnhGlnleuMetAspreuargnglnnaengln 2022  
QY 517 ----ATGAAAAGGAATAAGAAAGAGGAGTACATGGATCAAGATGTTG----- 567  
Db 2023 AlaLeuGlnlysglnleuGlnlysmetarglysrPheLeuaspcluglnlalaIleasprarg 2042  
QY 568 -----ATCTGTCTCAGAAATATGCCCCAACCTGAGGCCAGCTGGAAGA 612  
Db 2043 GlnhIsglnIuargAspralPheGlnGlnGlnlIleGlnlyseuGlnlGlnlneulysval 2062  
QY 613 GTTACAAGGAAAGATTTCACCTATTAATCAATGAGAAATTCAAAGCCAGTGGCT 672  
Db 2063 ValProArgPheGlnProIleSerGlnhIsglnThrArglnvalglu---GlnleuAla 2081  
QY 673 TCTCGGGGAATGGATGTCACAAAGTGTGGAGAAATG-----CGCTATCAGCTG 723  
Db 2082 AsnHISLeuLySGlnlysrPlyscysSerGlnleuLeuSerlysglnleu 2101  
QY 724 AATAAACCAACATGAGAGAGATGAGCAAGCAAAAGCACAGAGTTCAGAGCAAAA 783  
Db 2102 GlnArgAspIleGlnlulargasnlu---GlnIleGlnlyseuGlnrPheargValarg 2120  
QY 783 ----- 783  
Db 2121 GlnleuGlnlAlaLeuLeuGlnIuAsprarglysrPheGlyAlaValGlnAlaLysPro 2140  
QY 783 ----- 783  
Db 2141 GlnleuSerleuGlnValGlnleuGlnAlaGlnuArgAspAlaIleasprarglysglnlys 2160  
QY 784 -----ACTAAC-----AGGATCTGAAATTAAGAT 810  
Db 2161 GlnIleThrAsnleuGlnlulGlnleuGlnlGlnlPheargGlnlulGlnleuGlnlysnlysa 2180  
QY 811 CAGGAAATAGAGAAATGAGAAATAGAACTGAT-----GAAAGCAAAACAAAC 858  
Db 2181 GlnGlnIuValGlnlGlnIleuHISmetGlnleuGlnlIleGlnlysglnSerThrThrarg 2200  
QY 859 TTGGAAACAGAGACAGAGAGGCCCTGGCCAGAGAGAGTCCCTGAGACTAACAGAA 918  
Db 2201 LeuGlnGlnleuGlnlGlnlulasnlysrPheLyAspAspmetGlnlysrleuGlyLeu 2220  
QY 919 CTGCTGGCGCAATCTGAG-----CACCAACTGCACCTCACAGATCTGAATA 966  
Db 2221 AlaIleLyGlnSerAspralmetSerThrGlnAspGlnhISValleuPheGlyLysPhe 2240  
QY 967 GCTCAACTGATCAAGAAAGGATATACATATGATAAATGCGA-----AAG 1014  
Db 2241 AlaGlnIleIleIleGlnlulysGlnvalGlnIleAspGlnleuasnGlnlvalThrLys 2260  
QY 1015 TTACGAGAGAGA-----AATGAAGAAATTGAGAGAA----- 1044  
Db 2261 LeuGlnGlnlGlnleuLysIleThrThrAspAsnLysValIleleuGlnlysrasnGlnleu 2280  
QY 1045 -----CAGTGTCTCCAGCAT 1059  
Db 2281 IleArgAspLeuGlnlulThrGlnlIleGlnlysrleuMetSerAspGlnlulysValLysArg 2300  
QY 1060 GGGAGAGTACATGAAGCATGAAGCAAAAGCTTAAGGCACTGATTAAGCACAGCCAGGCC 1119  
Db 2301 AsnArg-----GlnlulGlnlIleGlnlulGlnleuasnGlnlvalIleGlnlysr 2315  
QY 1120 ACAGCCCAAGACGTGTCAGCTCCTCAGCAAGCAAGCAACCACTTCTCCTGAGAGAGCAG 1179

Db 2316 LeuGlnGlnlGlnleuAlaasnIleGlnlysrThrSer-----MetAsnAlaHIS 2332  
QY 1180 AGCCTGTGGAAGAGGTGGAGCCGCTGCGGACCACTTA 1218  
Db 2333 SerleuSerGlnlulAlaAspSerleuLysHISGlnleu 2345

Search completed: March 21, 2003, 13:52:22  
Job time : 90.5926 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 21, 2003, 12:44:22 ; Search time 164,803 Seconds  
(without alignments)  
12143.272 Million cell updates/sec

Title: US-09-502-945-1  
Perfect score: 2732  
Sequence: 1 cttctgcatgcacgcagaa.....aaatgaactttaagaaga 1552

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4569144 segs, 644733110 residues  
Total number of hits satisfying chosen parameters: 9138288

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL-frame+ n2p model -DEV=zip  
-Q/cgn2\_1/USPTO.spool/US09502945/runat\_14032003\_101100\_19180/app-query.fasta.1.10979  
-DB=Pending\_Patents\_AA\_Main -QMT=fastan -SUFFIX=rapm -MINMATCH=0.1 -LOOPEL=0  
-LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=US09502945 @cgn.1.1.1769 @runat.14032003\_101100\_19180 -NCPu=3  
-NO\_XLPRY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WAIN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: Pending\_Patents\_AA\_Main:\*  
2: /cgn2\_6/ptodata/1/paa/PCTUS\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/paa/US06\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/paa/US080\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/paa/US081\_COMB.pep:\*  
7: /cgn2\_6/ptodata/1/paa/US082\_COMB.pep:\*  
8: /cgn2\_6/ptodata/1/paa/US083\_COMB.pep:\*  
9: /cgn2\_6/ptodata/1/paa/US084\_COMB.pep:\*  
10: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep:\*  
11: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep:\*  
12: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep:\*  
13: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep:\*  
14: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep:\*  
15: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep:\*  
16: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep:\*  
17: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep:\*  
18: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep:\*  
19: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep:\*  
20: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep:\*  
21: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep:\*  
22: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep:\*  
23: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep:\*  
24: /cgn2\_6/ptodata/1/paa/US099\_COMB.pep:\*  
25: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep:\*  
26: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep:\*  
27: /cgn2\_6/ptodata/1/paa/US102\_COMB.pep:\*  
28: /cgn2\_6/ptodata/1/paa/US103\_COMB.pep:\*  
29: /cgn2\_6/ptodata/1/paa/US104\_COMB.pep:\*  
30: /cgn2\_6/ptodata/1/paa/US105\_COMB.pep:\*  
31: /cgn2\_6/ptodata/1/paa/US106\_COMB.pep:\*  
32: /cgn2\_6/ptodata/1/paa/US107\_COMB.pep:\*  
33: /cgn2\_6/ptodata/1/paa/US108\_COMB.pep:\*  
34: /cgn2\_6/ptodata/1/paa/US109\_COMB.pep:\*  
35: /cgn2\_6/ptodata/1/paa/US110\_COMB.pep:\*  
36: /cgn2\_6/ptodata/1/paa/US111\_COMB.pep:\*  
37: /cgn2\_6/ptodata/1/paa/US112\_COMB.pep:\*  
38: /cgn2\_6/ptodata/1/paa/US113\_COMB.pep:\*  
39: /cgn2\_6/ptodata/1/paa/US114\_COMB.pep:\*  
40: /cgn2\_6/ptodata/1/paa/US115\_COMB.pep:\*  
41: /cgn2\_6/ptodata/1/paa/US116\_COMB.pep:\*  
42: /cgn2\_6/ptodata/1/paa/US117\_COMB.pep:\*  
43: /cgn2\_6/ptodata/1/paa/US118\_COMB.pep:\*  
44: /cgn2\_6/ptodata/1/paa/US119\_COMB.pep:\*  
45: /cgn2\_6/ptodata/1/paa/US120\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1990.5	72.9	615	14	US-09-012-141-8
2	1990.5	72.9	615	19	US-09-502-967-70
3	1987.5	72.7	713	24	US-10-031-913-55
4	1296.5	47.5	355	1	PCT-US00-35017A-801
5	971	35.5	203	11	US-08-788-789-12
6	841	30.8	393	27	US-60-389-987-766
7	841	30.8	393	27	US-60-412-418-766
8	537	19.7	172	1	PCT-US01-03800A-1921
9	452	16.5	138	22	US-09-834-366-1401
10	452	16.5	138	27	US-60-197-873-14401
11	449	16.3	94	1	PCT-US02-09921-968
12	310	11.3	61	22	US-09-834-366-21271
13	310	11.3	61	27	US-60-197-873-21271
14	300	11.0	62	20	US-09-621-976-5141
15	300	11.0	62	27	US-60-147-499-5141
16	268	9.8	3899	25	US-10-171-311-4
17	268	9.8	3917	25	US-10-171-311-8
18	264	9.7	3878	25	US-09-614-259-11
19	264	9.7	3907	25	US-10-171-311-2
20	264	9.7	3911	27	US-60-389-987-1839
21	264	9.7	3911	27	US-60-412-418-1839
22	264	9.7	3925	25	US-10-171-311-6
23	262	9.6	860	24	US-10-037-417-59
24	262	9.6	860	24	US-10-080-334-166
25	256	9.4	1790	27	US-60-360-039-1586
26	255.5	9.4	576775	12	US-08-895-611-2
27	255.5	9.4	576775	18	US-08-895-611D-2
28	255.5	9.4	576775	12	US-09-458-180-2
29	255.5	9.4	576775	22	US-09-895-611D-2
30	254	9.3	2056	20	US-09-614-150-4824
31	254	9.3	2056	27	US-60-191-637-4824
32	254	9.3	2056	27	US-60-191-637-4824
33	253.5	9.3	3225	27	US-60-389-987-254
34	253.5	9.3	3225	27	US-60-412-418-254
35	253.5	9.3	3244	1	PCT-US01-14827-10091
36	251	9.2	1020	25	US-10-179-131-5887
37	251	9.2	1881	1	PCT-US02-03987-15590
38	251	9.2	1881	24	US-10-032-585-7646
39	251	9.2	1881	24	US-10-072-851-15590
40	251	9.2	1881	27	US-60-314-050-7646
41	249	9.1	1161	27	US-60-173-464-14216
42	249	9.1	1294	20	US-09-614-150-17298
43	249	9.1	1294	27	US-60-191-637-17298
44	249	9.1	1294	27	US-60-191-637-17298
45	246.5	9.0	1355	24	US-10-080-334-161

#### ALIGNMENTS

RESULT 1  
US-09-012-141-8  
Sequence 8, Application US/09012141  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: Lavallie, Edward R.  
APPLICANT: Racine, Lisa A.  
APPLICANT: Treacy, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Spaulding, Vikki  
APPLICANT: Agostino, Michael J.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
TITLE OF INVENTION: ENCODING THEM  
NUMBER OF SEQUENCES: 33

```

CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/012,141
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ. ID NO.: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 615 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-012-141-8

Alignment Scores:
Pred. No.: 1.79e-137 Length: 615
Score: 1990.50 Matches: 409
Percent Similarity: 83.81% Conservative: 0
Best Local Similarity: 83.81% Mismatches: 0
Query Match: 72.86% Indels: 79
DB: 14 Gaps: 1

US-09-502-945-1 (1-1552) x US-09-012-141-8 (1-615)
QY 16 GGAAGCTAAAGCTTACTATAGGAAAGTGAATGAGAAATCCCAATTGAAGTTT 75
DB 128 GLLTSLLeuLysLeuThrYrGluGluLysCysGluLeuLysSerGlnLeuLysS 147
QY 76 TTGAGAACGACTTAGCTGAATATCAGAGAACTTGAAGATCTTAAGAGCACTAAAG 135
DB 148 LeuArgAsnAspLeuLeuLeuGluYrGlnArgThrCysGluAspLeuLysGlnLeuLys 167
QY 136 CATTAAGAAATTTCTTGGCTGCTAATACTTGAACCGTGTGGTCTTTGTTGAAA 195
DB 168 HSLYSGLIuPheLeuLeuLalaIAsnThrCysAsnArgValGlyGlyLeuLysLeuLys 187
QY 196 TGTGGCTCAGCATGAGCTGTCTTCCCAACCCCACTAATCTTATATGAGACAC 255
DB 188 CysAlaGlnHisGluLalaValLeuSerGlnThrHisThrAsnValHisMetGlnThrIle 207
QY 256 GAAAGACTGTTAAAGAAAGAGATGACTGATGCTGCATAGTTCGCTAAGAGACAC 315
DB 208 GLuArgLeuValLysGluArgAspAspLeuMetSerAlaLeuValSerValArgSerSer 227
QY 316 TTGGCAGATACGACGAAAGAGAACGAACTGCTTATGAACAGGTGAACAAAGTTTGC 375
DB 228 LeuAlaAspThrGlnGlnArgGlnAlaSerAlaTyrgLysGlnValLysGlnValLeuGln 247
QY 376 AATATCGAGAGAGCCAAATTTTGAAGAAACCAAGGCTTTATATCCAGTGTACCAAGTTGAG 435
DB 248 ILeSerGluGlnLalaAsnHegLysThrLysAlaLeuLysGlnLysAspGlnLeuArg 267
QY 436 AAGGAGCTGAGAGGACGCGAGGACGACTTGAAGAAAGAACTTGATCAGCAAGAGAAA 495
DB 268 LysGluLeuGlnArgGlnAlaGluArgLeuGluLysGluLeuAlaSerGlnGlnGlnLys 287

```

```

QY 496 AGGCCATTGAGAAAGACATGATGAAAAAGAAATACGAAAGAAAGGAGTCAATGGGA 555
DB 288 ArgAlaIleGluLysAspMetLysGluIleThrLysGluArgGluTyrMetGly 307
QY 556 TCAAGATGTTGATCTTGTCTGCAATATGCTCCCACTGGAGGCCAGGAGGAAAGGTT 615
DB 308 SerLysMetLeuLysSerGlnAsnIleAlaGlnLeuGluAlaGlnValGluLysVal 327
QY 616 ACAAGGAAAGATTTCACTATTAATCACTGAGAGAAATTCAGAGCCAGCTGGCTTC 675
DB 328 ThrLysGluLysIleSerAlaIleAsnGlnLeuGlnGluIleGlnSerGlnLeuAlaSer 347
QY 676 CGGGAATGGATGTCACAAAGCTGTGTGGAAGAAATGCGCTATAGCTGAATAAACCAAC 735
DB 348 ArgGluMetAspValThrLysValCysGlyGluMetArgTyrGlnLeuAsnLysThrAsn 367
QY 736 ATGAGAGATGATGACAGAAAGAGGACGACGAGCTTCAGACCAAACTAACAGGAT 795
DB 368 MetGluLysAspGluAlaGluLysGluHisArgGluPheArgAlaLysThrAsnArgAsp 387
QY 796 CTTGAATTAAGATGACGAAATAGAGAAATGAGAATGAACTGGATGAACCAACAA 855
DB 388 LeuGluIleLysAspGlnGluIleGluLysLeuArgIleGluLeuAspGluSerLysGln 407
QY 856 CACTTGAACAGGACGACGAGAGAGGACCCCTGGCCAGAGAGAGTGTGAGACTTAACA 915
DB 408 HisLeuGluGlnGlnGlnGlnLysAlaAlaLeuAlaArgGluGluLysLeuArgLeuThr 427
QY 916 GAATCTGCTGGCGCAATCTAGCCACCACTGCACCTC----- 951
DB 428 GluLeuLeuGluGluSerGlnHisGlnLeuHisLeuThrArgGlnGluLysAspSerIle 447
QY 951 ----- 951
DB 468 GluLeuThrGlnLysIleGlnGlnMetGluAlaGlnHisAspLysThrGluAsnGluGln 487
QY 951 ----- 951
DB 488 TyrLeuLeuLeuThrSerGlnAsnThrPheLeuThrLysLeuLysGluLysCysThr 507
QY 952 ----- ACCAGATCTGAAATAGCTCAACTCAGT 978
DB 508 LeuAlaLysLysLeuGluGlnLysSerGlnLysThrArgSerGluIleAlaGlnLeuSer 527
QY 979 CAAGAAAAAGGATATCATATGATAATTTGGGAAAGTTACAGAGAAAGAAATGAAGATTG 1038
DB 528 GlnGluLysArgTyrThrTyrAspLysLeuGluLysLeuGlnArgArgAsnGlnGluLeu 547
QY 1039 GAGGACAGTGTGTCCAGCATGAGAGTACATGACATGACATGAGCAAGGCTTAAGGACG 1098
DB 548 GlnGluGlnCysValGlnHisGlyArgValHisGluThrMetLysGlnArgLeuAlaGln 567
QY 1099 CTGGATTAACACACAGCCAGGACGACCCAGAGCTGTGCTGAGCTCTCCAGACAGCAAC 1158
DB 568 LeuAspLysHisSerGlnAlaThrAlaGlnGlnLeuValGlnLeuLeuSerLysGlnAsn 587
QY 1159 CAGCTTCTCTGAGAGGACGACAGACCTGTGGAAGAGGTGACCGGCTGGAGCCAGTTA 1218
DB 588 GlnLeuLeuLeuGlnLysGlnSerLeuSerGlnGluValAspArgLeuArgThrGlnLeu 607
QY 1219 CCCAGATGCCCAATCTGATTGC 1242
DB 608 ProSerMetProGlnSerAspCys 615

RESULT 2
US-09-502-967-70
; Sequence 70, Application US/09502967
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth

```

```

: APPLICANT: McCoy, John M.
: APPLICANT: Lavallee, Edward R.
: APPLICANT: Collins-Racie, Lisa A.
: APPLICANT: Evans, Cheryl
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Spaulding, Vikki
: APPLICANT: Genetics Institute, Inc.
: TITLE OR INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
: FILE REFERENCE: GI 6021-30X
: CURRENT APPLICATION NUMBER: US/09/502,967
: CURRENT FILING DATE: 2000-02-11
: NUMBER OF SEQ ID NOS: 306
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 70
: LENGTH: 615
: TYPE: PR1
: ORGANISM: Homo sapiens
US-09-502-967-70

```

```

Alignment Scores:
Pred. No.: 1,79e-137 Length: 615
Score: 1990.50 Matches: 409
Percent Similarity: 83.81% Conservative: 0
Best Local Similarity: 83.81% Mismatches: 0
Query Match: 72.86% Indels: 79
DB: 19 Gaps: 1
US-09-502-945-1 (1-1552) x US-09-502-967-70 (1-615)

```

```

QY 16 GAGAGCTAAACCTTACTTATGAGAAAGTGTGAATTTGAGAAATCCCAATGAGTTT 75
DB 128 Gtutysleuylsleuthrtytgytuglulyscysglutlleglulsergluleuylshe 147
QY 76 TTGAGCAACGACTTACGTAATATCAGAGAACTGTGAAGATCTTAAAGCAACTAAAG 135
DB 148 leuairgsnaaspleuadlaglwtglnarthyrcysgluaspuleuylsglulnleuyls 167
QY 136 CATTAAGAAATTTCTTCGCGCTAATCTGTAAACCGTGTGGTGGCTTTGTTGAAA 195
DB 168 HlslsyglupheleuLeuAlaAlasnThrcysasnArgvalGlyglYleucysleuyls 187
QY 196 TGTGCTGAGCATGAGAGCTGTTCTTCCCAACCCATACTATGTTCATATGCAGACATC 255
DB 188 CyslaaglnhlsgluAlaValleuSerclnThrHsthrahnValHlsMeclnThrlle 207
QY 256 GAAAGACTGTTAAAGAAAGATGACTTGATGTCGCACTAGTTTCCGTAAAGAGCAGC 315
DB 208 GluArgleuVallylsGluArGAspAspleuMetSerAlaValSerValArgSerSer 227
QY 316 TTGGCAGATTGCGCAAGAGAAAGCAAGTCTTATGAACAGGTGAACAAGTTTTCGAA 375
DB 228 leuAlasprlnrglnGlnArgGluAlaSerAlaTyrglnuglnValylsGlnValleuGln 247
QY 376 ATATCTGAGAAAGCAATTTTGAAGAAAAACCAAGCTTTAATCCAGTGTGACCAAGTGAAG 435
DB 248 lleserGlnGluAlaAsnPhcglulserThrylsAlaLeuIlleglnCysAspGlnleuArg 267
QY 436 AAGGAGCTGGAGAGCGAGCGGAGCGACTTGAAAAAGAACTTGACATCTCAGCAAGAGAA 495
DB 268 lylsGlnleuGlnuArglnAlaGlnArgleuGlnulysGlnleuAlaSerGlnGlnGln 287
QY 496 AGGGCCATTGGAAGAGACATGATGAAAAAGAAATTAAGGAAAGAAAGGAGATACATGGGA 555
DB 288 ArgAlaIleGlnuylsAspMetMetlylsGlnlletHrlylsGlnuArgGlnuTyMetlY 307
QY 556 TCAAGAGATGTGATCTGTCTCAGAAATATTCGCCAACGTGAGGCGCCAGGTGGAAGAGTT 615
DB 308 SerlyshetleuIlleuSerGlnAsnIlleAlaGlnleuGlnuAlaGlnValGlnulysVal 327
QY 616 ACAAGAGAAAGATTTGAGCTATTATCACTGAGAGAAATTCAAAGCCAGCTGCTTCT 675
DB 328 ThrlysglnuylsIleserAlaIleAsnGlnleuGlnuIlleglnInserGlnleuAlaSer 347

```

```

QY 676 CGGAAATGATGTCTACAAGAGTGTGTGAGAAATGCCCTATACGCTGAATAAAAACCAAC 735
DB 348 ArgGlnMetAspValThrllysValCysGlyGlnMetArgTyrglnleuAsnlysthrahn 367
QY 736 ATGGAAGAGATGAGGCGAAAGAGACAGAGAGTTCAGAGCAAAACTAACAGGAT 795
DB 368 MetGlnlyAspGlnuAlaGlnulysGlnHlsArgGlnPheArgAlaTysthrahnArgAsp 387
QY 796 CTTGAAATTAAGATCAGAGAAATAGAGAAATTTGAGAAATGAACTGGATGAAGCAAAACA 855
DB 388 leuGlnuIlleuylsAspGlnGlnuIlleglnuylsleuArGylleglnleuAspGlnSerlysln 407
QY 856 CACTTGGACAGAGACACAGACAGAGAGCCCTGGCCAAAGAGAGAGTGCCTAGACTACA 915
DB 408 HlsleuGlnGlnGlnGlnGlnGlnlyslAlaLeuAlaArgGlnGlnCysleuArGlnleuThr 427
QY 916 GAAGTGTGGGCGAATCTGAGCACCACCACTGCACCTC----- 951
DB 428 GlnleuLeuGlnGlnSerGlnHlsGlnleuHlsleuThrArgGlnGlnulysAspSerlle 447
QY 951 ----- 951
DB 468 GlnleuThrGlnulysIlleglnGlnMetGlnuAlaGlnHlsAspLysthGlnuAsnGlnGln 487
QY 951 ----- 951
DB 488 TyrlleuLeuLeuThrSerGlnAsnThrPheleuThrlyslsleuylsGlnGlnCysCysThr 507
QY 952 -----ACCAATGTGAATATGCTCAACTAGT 978
DB 508 leuAlaTyrlsleuGlnGlnIlleserGlnlysthThrArgSerGlnuIlleAlaGlnleuSer 527
QY 979 CAAGAAAGAGTTATACATATGATTAATTTGGAAAGTTACAGAGAGAAAGAAAGAAATG 1038
DB 528 GlnGlnuylsArgTyrlThrTyrlAspLyslsleuGlnuArgAsnGlnGlnleu 547
QY 1039 GAGGAACAGTGTGTCCACATGAGGAGATGATGAGATGAAGCAAGGCTTAAGCGAG 1098
DB 548 GlnGlnGlnCysValGlnHlsGlyArgValHlsGlnuThrMetlyslsGlnuArgleuArgln 567
QY 1099 CTGGATTAAGCACAGCCAGCCAGCCAGCAAGCTGTGCAAGCTCTAGCAAGCAGAAC 1158
DB 568 leuAspLyshlsSerGlnAlaThrlaGlnGlnleuValGlnleuLeuSerlyslsGlnAsn 587
QY 1159 CAGCTTCTCTGAGAGAGCAGAGGCTGTGCGAAAGAGTGGAGCCGCTCGGACCCAGTTA 1218
DB 588 GlnleuLeuLeuGlnuArglnSerleuSerGlnuValAspArgleuArgThrlGlnleu 607
QY 1219 CCCAGCATGCCACAATCTGATTC 1242
DB 608 ProserMetProGlnSerAspCys 615

```

```

RESULT 3
US-10-031-915-52
: Sequence 52, Application us/10031915
: GENERAL INFORMATION:
: APPLICANT: INCYTE GENOMICS, INC.
: APPLICANT: HILLMAN, Jennifer L.
: APPLICANT: LAL, Preeti
: APPLICANT: TANG, Y. Tom
: APPLICANT: YUE, Henry
: APPLICANT: AU-YOUNG, Janice
: APPLICANT: BANDMAN, Olga
: APPLICANT: AZIMZAI, Yalda
: APPLICANT: YANG, Junning
: APPLICANT: LU, Dying Alpha M.
: APPLICANT: BAUGHN, Mariah R.
: APPLICANT: PATTERSON, Chandra

```

```

; APPLICANT: SHAH, Purvi
; TITLE OF INVENTION: CELL CYCLE AND PROLIFERATION PROTEINS
; FILE REFERENCE: PF-0722 PCT
; CURRENT APPLICATION NUMBER: US/10/031, 915
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: 60/145,075; 60/153,129; 60/164,647
; PRIOR FILING DATE: 1999-07-21; 1999-09-08; 1999-11-10
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PERL Program
; SEQ ID NO: 52
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 5678487CD1
US-10-031-915-52

```

```

Alignment Scores:
Pred. No.: 3,05e-137 Length: 713
Score: 1987.50 Matches: 408
Percent Similarity: 83.81% Conservative: 1
Best Local Similarity: 83.61% Mismatches: 0
Query Match: 72.75% Indels: 79
DB: 24 Gaps: 1

```

US-09-502-945-1 (1-1552) x US-10-031-915-52 (1-713)

```

QY 16 GAAAGCTAAAGCTTCTTGTGAGAAAGTGGAAATTCGAATCCCAATGGAAGTTT 75
DB 226 GLLuLyLeuLysLeuThrYrGlnGluLysCysGluLeuGlnSerGlnLeuLysPhe 245
QY 76 TTGAGACAGCTTACCTGATATCAGAGAACTTGGAAGATCTTAAAGCAACATAAG 135
DB 246 LeuAlaAsnAspLeuAlaGlnArgTrgTrgCysGlnAspLeuLysGlnLeuLys 265
QY 136 CATAAAGATTTCTTGGCTGCTAATCTGTGTAACCGTGTGGTCTTTGTGGAAA 195
DB 266 HisLysGlnPheLeuLeuAlaAlaAsnThrCysAsnArgValGlyLeuLysCysLeuLys 285
QY 196 TGGCTCAGCATGAGCTGTCTTCCCAAGCCATCAATGATTCATATGACAGCATC 255
DB 286 CysAlaGlnHisGlnAlaValLeuSerGlnThrHisThrAsnValHisMetGlnThrIle 305
QY 256 GAAAGACTGGTTAAAGAAAGATGACTGTGCTGCTAGCTTCCGTAAGAGAGAGC 315
DB 306 GlnArgLeuValLysGlnArgAspAspLeuMetSerAlaLeuValSerValArgSerSer 325
QY 316 TTGGCAGATACGAGCAAGAAAGCAAGCTTATGACAGGTGAACAAAGTTTGGCAA 375
DB 326 LeuAlaAspThrGlnGlnArgGlnAlaSerAlaTrgGlnValLysGlnValLeuGln 345
QY 376 ATATCTGAGGAAGCCAAATTTTGAAGAAACCAAGGCTTTAATCCAGTGTGACAGTTGAG 435
DB 346 IleSerGlnGlnAlaAsnPheGlnLysThrLysAlaLeuIleGlnCysAspGlnLeuArg 365
QY 436 AAGAGCTGGAGAGGCGGAGCGAGCACTTGAAAGAAAGCTTCATCTCAGCAAGAGAAA 495
DB 366 LysGlnLeuGlnLysArgGlnAlaArgLeuGlnLysAspLeuAlaSerGlnGlnLys 385
QY 496 AGGGCCATTCAGAAAGACATGATGAAAGAAAGAAATACGAAAGAGGAGTCAATGGGA 555
DB 386 ArgAlaIleLeuLysAspMetLysLysGlnIleThrLysGlnArgGlnLysGly 405
QY 556 TCAAGATGTTGATCTGTCTCAGATATATGCCCACATGGAGGCCAGGTGAGAAAGGTT 615
DB 406 SerLysMetLeuIleLeuSerGlnAsnIleAlaGlnLeuGlnAlaGlnValGlnLysVal 425
QY 616 ACAAGAGAAAGATTCAGCTATTAAATCAACTGAGAGAAATTCAAAGCCAGCTGCTCT 675
DB 426 ThrLysGlnLysIleSerAlaIleAsnGlnLeuGlnLysGlnIleGlnSerGlnLeuAlaSer 445
QY 676 CGGGAATGGATGTCAAAAGGTGTGTGGAGAAATGCGCTATCAAGCTGAATAAACCAAC 735

```

```

DB 446 ArgGlnMetAspValThrLysValCysGlyGlnMetArgTrgGlnLeuAsnLysThrAsn 465
QY 736 ATGAGAAAGATATGAGCAGAGAAAGAGACACAGAGAGTTCAGACCAAAACTAACAGGAT 795
DB 466 MetGlnLysAspLysAlaGlnLysGlnHisArgGlnPheArgAlaLysThrAsnArgAsp 485
QY 796 CTGGAATTAAGATCAGAAATAGAGAAATTCAGAAATGCAATGATGCAAGCAACAA 855
DB 486 LeuGlnIleLysAspGlnGlnIleGlnLysLeuAlaArgIleGlnLeuAspLysSerLysGln 505
QY 856 CACTTGAAACAGAGACAGCAGAAAGCAAGCCCTGGCCACAGAGAGAGTCCCTGAGACTTACA 915
DB 506 HisLeuGlnGlnGlnGlnGlnLysAlaAlaLeuAlaArgGlnLysLeuArgLeuThr 525
QY 916 GAAGCTGTGGGCAATCTGAGCAACCAACTGCACCTC----- 951
DB 526 GlnLeuLeuGlnGlnSerGlnHisGlnLeuHisLeuThrArgGlnLysAspSerIle 545
QY 951 ----- 951
DB 546 GlnGlnSerPheSerLysGlnAlaLysAlaGlnAlaLeuGlnAlaGlnArgGlnGln 565
QY 951 ----- 951
DB 566 GlnLeuThrGlnLysIleGlnGlnMetGlnAlaGlnHisAspLysThrGlnAsnGlnGln 585
QY 951 ----- 951
DB 586 TyrLeuLeuLeuThrSerGlnAsnThrPheLeuThrLysLeuLysGlnLysCysThr 605
QY 952 -----ACAGATCTGAATAGCTCAACTCAGT 978
DB 606 LeuAlaLysLysLeuGlnGlnIleSerGlnLysTrgArgSerGlnIleAlaGlnLeuSer 625
QY 979 CAAGAAAAAGGATATCATATGATTAATTTGGAAAGTTACAGAGAAATGAAGATTTG 1038
DB 626 GlnGlnLysArgTrgTrgTrgAspLysLeuGlnLysLeuGlnArgAsnGlnGlnLeu 645
QY 1039 GAGGACAGGTGTCTCCAGCATGGAGAGTACATGAGACAGTGAAGCAAAAGCTTAAGCAG 1098
DB 646 GlnGlnGlnCysValGlnHisGlnLysArgValHisGlnThrMetLysGlnArgLeuArgGln 665
QY 1099 CTGGATTAAGCAGACGAGCCAGCCAGCCAGCTGTGTGAGCTCTCAGCAAGCAGAAC 1158
DB 666 LeuAspLysHisSerGlnAlaThrAlaGlnGlnLeuValGlnLeuLeuSerLysGlnAsn 685
QY 1159 CAGCTTCTCTGGAGAGGCGAGCGCTGTGGAGAGGTGAGCGGCTGGGAGCCAGTTA 1218
DB 686 GlnLeuLeuLeuGlnLysGlnSerLeuSerGlnLysValAspArgLeuArgTrgGlnLeu 705
QY 1219 CCCAGCATGCCACAAATCTGATTGC 1242
DB 706 ProSerMetProGlnSerAspCys 713

RESULT 4
PCT-US00-35017A-801
; Sequence 801, Application PC/TUS0035017A
; GENERAL INFORMATION:
; APPLICANT: Hysed Inc
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: PCT/US00/35017A
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 801
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Homo sapiens

```

PCT-US00-35017A-801

## Alignment Scores:

Pred. No.:	1.7e-86	Length:	355
Score:	1296.50	Matches:	274
Percent Similarity:	77.62%	Conservative:	0
Best Local Similarity:	77.62%	Mismatches:	0
Query Match:	47.46%	Indels:	79
	1	Gaps:	1

US-09-502-945-1 (1-1552) x PCT-US00-35017A-801 (1-355)

```

QY 244 ATGCAGACATCGAAGACTGGTTAAGAAAAGAGATGATGTCGACTGATTC 303
    |||
Db 1 MetGlnThrIleGlnIuArgLeuValIySGluArgspAspLeuMetSerAlaLeuValSer 20
QY 304 GTAAGGAGCAGCTTGGAGATACGACAGCAAGACAGACAGCTTATGAAACAGTGA 363
    |||
Db 21 ValArgSerSerLeuAlaAspThrGlnGlnArgIuAlaSerAlaTYrGlnIuValIys 40
QY 364 CAACTTTGCAATATCTGAGAGAGCCAAATTTGAAAAACCAAGCGCTTAATCCAGT 423
    |||
Db 41 GlnValLeuGlnIleSerGluGluAlaAspPheGluIuIuSerIlySalAlaLeuIleGlnCys 60
QY 424 GACCACTTGAAGAGAGAGCTGAGAGGCGAGCGAGGAGCTTGAAGAAAGACTTGACATCT 483
    |||
Db 61 AspIleuArgIySGluLeuGlnArgGlnAlaGluArgLeuGluIySGluLeuAlaSer 80
QY 484 CAGCAAGAGAAAAGGCCATTGAGAAAGACATGATGAAAAAGAAATACCAAGAAAG 543
    |||
Db 81 GlnGlnIuIySArgAlaIleGluIySAspMetMetIySGluIleThrIySGluArg 100
QY 544 GAGTACATGGGATCAAGATGTTGATCTGTCAGAAATATGGCCCAACTGAGAGCCAG 603
    |||
Db 101 GluIyMetGlySerIySMeLeuIleLeuSerGlnAsnIleAlaGlnLeuGluAlaGln 120
QY 604 GTGGAAGAGGTTACAAAGAGAAAGATTTCAGCTTATATCACTGAGAGAAATCAAGC 663
    |||
Db 121 ValGluIyValIThrIySGluIySleSerAlaIleAsnGlnLeuGluIleGlnSer 140
QY 664 CAGCTGGCTTCTCGGAAATGATGTCAAAAGGTGTGTGAGAAATGCCATCAGCTG 723
    |||
Db 141 GlnLeuAlaSerArgGluMetAspValIThrIyValCysGluMetArgTYrGlnLeu 160
QY 724 AATTAACCAACATGAGAGAGATGAGCAAGAAAGAGACAGACAGAGCTTCAGAGCAAAA 783
    |||
Db 161 AsnIySThrAsnMetGluIySAspGluAlaGluIySGluIuIuAspArgIuAlaIys 180
QY 784 ACTAACAGGATCTTGAATTAAGATCAGAAATAGAGAAATGAGAATAGACTGGAT 843
    |||
Db 181 ThrAsnArgspLeuGluIleIySAspGlnIuIleGluIySLeuArgIleGluLeuAsp 200
QY 844 GAAAGCAACACACTTGAAGACAGAGACAGACAGAGAGCCCTGGCCAGAGAGAGTGC 903
    |||
Db 201 GluSerIySGlnHisLeuGlnGlnGlnGlnGlnGlnIySalAlaLeuAlaArgGluIyCys 220
QY 904 CTGAGATACAGACACTGCTGGGGGATCTGAGACCAACATGCACCTC----- 951
    |||
Db 221 LeuArgLeuThrGluLeuLeuGluIyGlnSerGlnHisGlnLeuHisLeuThrArgGlnGln 240
QY 951 ----- 951
Db 241 LysAspSerIleGlnGlnSerPheSerIySGluAlaIySalGlnAlaLeuGlnAlaGln 260
QY 951 ----- 951
Db 261 GlnArgGluGlnGluLeuThrGlnIySleGlnGlnMetGluAlaGlnHisAspIySThr 280
QY 951 ----- 951
Db 281 GluAsnGlnGluIyIyLeuLeuLeuThrSerGlnAsnThrPheLeuThrIySLeuIySGlu 300
QY 952 -----ACCAATCTGGAATA 966
  
```

```

Db 301 GluCysCysThrIleuAlaIySLeuGluGlnIleSerGlnIySThrArgSerGluIle 320
QY 967 GCTCACTCACTCAAGAAAAAGTATACATATATAATTGGAAAGTTACAGAGAGA 1026
    |||
Db 321 AlaGlnLeuSerGlnGlnIyIySArgTYrThrTyrsPysIySGluIySLeuGlnArg 340
QY 1027 AATGAGAATTTGAGAGACAGTGTCTCAGCATGAGAGA 1065
    |||
Db 341 AsnGluGluLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 353

RESULT 5
US-08-788-789-12
: Sequence 12, Application US/08788789
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John
: APPLICANT: Lavalley, Edward
: APPLICANT: Racie, Lisa
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Spaulding, Vikki
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 Cambridgepark Drive
: CITY: Cambridge
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/788,789
: FILING DATE:
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Brown, Scott A.
: REGISTRATION NUMBER: 32,724
: TELEPHONE: (617) 876-8224
: TELEFAX: (617) 876-5851
: INFORMATION FOR SEO ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 203 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-788-789-12

Alignment Scores:
Pred. No.: 1.31e-62 Length: 203
Score: 971.00 Matches: 201
Percent Similarity: 99.01% Conservative: 0
Best Local Similarity: 99.01% Mismatches: 2
Query Match: 35.54% Indels: 0
DB: 11 Gaps: 0

US-09-502-945-1 (1-1552) x US-08-788-789-12 (1-203)
QY 286 ATGCTCAGTCACTGTTCCGTAAGAGAGAGCTTGGCAGATACGAGCAAGAGAGAGT 345
    |||
Db 1 MetSerAlaLeuValSerValArgSerSerLeuAlaAspThrGlnGlnAlaIaSer 20
QY 346 GCTTATGACAGAGGTGAAGCAAGTTTGCATAATCTGAGAGAGCAATTTTGAAGAAACC 405
    |||
Db 21 AlaTyrgluGlnValIySGlnValIleuGlnIleSerGlnGlnAlaAsnPheGluIySThr 40
  
```



OY	406	AAAGCTTTAATCCAGTGTGACCACTTGTAGGAAGAGCGTGGAGAGCGGAGGCCACTT	465
Db	41	LysAlaLeuIleIleGlnCysAspPheLeuArgIleGlySLeuGlnGluArgGlnAlaIleGluArgLeu	60
OY	466	GAAGAAAGACTTGGCATCTCCACAGAGAGAGAAAGGGCCATTGTGAAAGACATGATGAAAG	525
Db	61	GluIysGluLeuAlaSerGlnGlnGlnIleuArgAlaIleGluIlysAspMetMetLysLys	80
OY	526	GAATTAACGAAAGAAAGGAGTACATGGAGTCAAAAGATGTGGATCTGTCTCAGAAATTT	585
Db	81	GluIleThrLysGluArgGluIutyMetCylSerLysMetLeuIleLeuSerGlnAsnIle	100
OY	586	GCCCAACTGAGAGGCCCAAGGTGGAAGAGTTTCAAAAGAGAAAGATTTCAGCTATTAACTAA	645
Db	101	AlaGlnLeuGluAlaGlnValGlnLysValThrLysGlyLysIleSerAlaIleAsnGln	120
OY	646	CTGAGAGAAATTCAAAGCCAGCTGGCTTCTCGGAATTCGATGTCACAAAGGTGTGGGA	705
Db	121	LeuGlnGluIleIleGlnSerGlnLeuAlaSerArgIleMetAspValThrLysValCysGly	140
OY	706	GAATGCGCTTACAGCTCAATATAAACCAACATCTGGAGAGGTGAGCGAGAGAAAGAGACAC	765
Db	141	GluMetArgItyrGlnLeuAsnLysThrAsnMetCylLysAspGluAlaGluLysGlnHis	160
OY	766	AGAGAGTTCAAGCCAAAACCTAACAGGGATCTTGAATTTAAAGATCCAGAAATAGAGAA	825
Db	161	ArgGluPheArgAlaLysThrAsnArgAsp**GluIleLysAspGlnGluIleGluLys	180
OY	826	TTGAGATTAAGAACTGGATGAAGAAGCAACACTTTGGAAACGAGAGCAGAGAAAGCAGCC	885
Db	181	LeuArgIleGluLeuAspGluSerLysGlnHisIleuGlnGlnGlnGlnLysAlaAla	200
OY	886	CTGGCCACA 894	
Db	201	LeuAlaArg 203	

```

RESULT 6
US-60-389-987-766
; Sequence 766, Application US/60389987
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Boi D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660008.465P2
; CURRENT APPLICATION NUMBER: US/60/389,987
; CURRENT FILING DATE: 2002-06-17
; NUMBER OF SEQ ID NOS: 3025
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 766
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-389-987-766

```

PragmEd. Locals:	
Pred. No.:	5.27e-53
Score:	841.00
Percent Similarity:	100.00%
Best local Similarity:	100.00%
Query Match:	30.78%
DB:	27
Length:	393
Matches:	168
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-09-502-945-1 (1-1552) x US-60-389-987-766 (1-393)

Qy 16 GAGAGCTAAAACCTTACTTATGAGGAAAGCTGAAATTGAGGAATCCCAATTGAAGTTT 75  
|||||

Db	226	GLIuylsleuLysleuThrTyrGlulglulLysScylulileglulgluSerGlnleuLysPhe	245
Qy	76	TTGAGCAACGACTTAACTGAAATATCAGAGAACTTGTAGACTCTTAAGAACACTAAAG	135
Db	246	LeuATgAaAsnAspLeuAlaGluTyrGluIArgThrCysGluAspLeuLysGluGlnleuLys	265
Qy	136	CATAAAGAATTCTTCTGCGTCGTAAATCTGTAAACCGTGTGGTGCCTTTGTTGAAA	195
Db	266	HisLysGluIleuLeuLeuAlaIAsnThrCysAsnArgAlaGlyGlyLeuCysleuLys	285
Qy	186	TGTGCTCAGCATGGAAGCTGTTCTTCCCAACCCATCTAATGTTCAATGCAAGACCATC	255
Db	286	CysAlaGlnHisGluAlaValAlaLeuSerGlnThrHisThrAsnValHisMetGlnThrIle	305
Qy	256	CAAAACACGCTTAAAGAAGAAGATGACTGATGTCTGCACAGTTCCTCGTAAGAGACAGC	315
Db	306	GluATgLeuValLysGluIArgAspIAspLeuMetSerAlaLeuValSerValArgSerSer	325
Qy	316	TTGCGCAGATACGAGCAAGAAGAAGCAAGTGCCTTATCAACAGCGAAGCAAGTTTGCAA	375
Db	326	LeuAlaAspTrpGlnGlnIArgGluAlaSerAlaTyrGluGlnValLysGlnValleuGln	345
Qy	376	ATATCTGGAGGAAGCCAAATTTTGAAGAAAAACAAGCTTTAATCCAGTGTGACCAAGTTGAGG	435
Db	346	IleSerGluGluAlaAsnPheGluLysThrLysAlaLeuIleGlnCysAspGlnleuArg	365
Qy	436	AAGGAGCTGGAGAGCGAGCGACGAGCTTGAAGAAAGCACTTGCGATTCAGCAAGAGAAA	495
Db	366	LysGluLeuGluIArgGlnAlaGluIArgleuLulysGluIAsnSerGlnGlnLys	385
Qy	496	AGGCGCATTTGAGAAAGACATGATG	519
Db	386	ArgAlaIleGluLysAspMetMet	393

```

RESULT 7
US-60-412-418-766
; Sequence 766, Application US/60412418
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TANGERS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 6600088_465P3
; CURRENT APPLICATION NUMBER: US/60/412,418
; CURRENT FILING DATE: 2002-09-20
; NUMBER OF SEQ ID NOS: 3025
; SOFTWARE: FASTSEQ For Windows Version 4.0
; SEQ ID NO 766
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-412-418-766

```

Alignment Scores:	
Pred. No.:	5,27e-53
Score:	841.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	30.78%
DB:	27
Length:	393
Matches:	168
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-09-502-945-1 (1-1552) x US-60-412-418-766 (1-393)

07 16 GAGAGCTTAAACTTACTTATGAGGAGAAAAGTGTAATTTGAGCAATCCCAATTGAAGTTT 75  
 |||||  
 07 226 GIuYstleuLYsleuThTYrGIuGIuLYsCseIuIeGIuGIuSerGIuLeuLYsPhe 245  
 |||||  
 07 76 TTGAGCAAGCACTTACGCTGAATATACAGAGAACTTTGTGAGACTTCCTTAAAGACAACTAAAG 135

|||||  
Db 246 LeuArGaNaSpLeuAlaGluTyrGlnArgThrCysGluAspLeuTyrGlnLeuLys 265  
OY 136 CATAAAGAAATTTCTTGGCTGCTTAATCTGTAAACCGTGTGGTCTTTGTTGAA 195  
Db 266 HisTyrGlnPheLeuLeuAlaAlaSerThrCysAsnArgValGlyGlyLeuCysLeuLys 285  
OY 196 TGTGTCAGCATGAAGCTGTCTTCCCAACCCATTAATGTTCATATGACAGACATC 255  
Db 286 CysAlaGlnHisGluAlaValLeuSerGlnThrHisThrAsnValHisMetGlnThrIle 305  
OY 256 GAAAGACTGTGTTAAAGAAAGAGATGACTGTGCTGCATAGTTTCCGTAAGAGCAGC 315  
Db 306 GluArgLeuValLysGluArgAspAspLeuMetSerAlaLeuValSerValArgSerSer 325  
OY 316 TTGGCAGATACGACCAAGAAAGCAAGCTGTATGACAGGTGAACAAGTTTGCAA 375  
Db 326 LeuAlaAspThrGlnArgGlnArgGlnAlaSerAlaTyrGlnGlnValLysGlnValLeuGln 345  
OY 376 ATATCTGAGGAAGCCATTTTGAAAAACCAAGCTTTAATCCAGTGTGACCATGTGAG 435  
Db 346 IleSerGlnGluAlaAsnPheGlnLysThrLysAlaLeuIleGlnCysAspLeuLysArg 365  
OY 436 AAGGAGCTGAGAGGCGAGCGAGCGACTGTGAAAAAGAACTTGATCTCAGCAAGAGAAA 495  
Db 366 LysGluLeuGlnArgGlnAlaGlnArgLeuGlnLysGlnLeuAlaSerGlnGlnLys 385  
OY 496 AGGCCATTTGAGAAAGACATGATG 519  
Db 386 ArgAlaIleGlnLysAspMetMet 393

## RESULT 8

PCT-US01-03800A-1921  
Sequence 1921, Application PC/TUS0103800A  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
APPLICANT: Tang, Y, Tom et al  
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides  
FILE REFERENCE: 21272-030  
CURRENT APPLICATION NUMBER: PCT/US01/03800A  
CURRENT FILING DATE: 2001-02-05  
PRIORITY APPLICATION NUMBER: 09/560,875  
PRIORITY FILING DATE: 2000-04-27  
PRIORITY APPLICATION NUMBER: 09/496,914  
PRIORITY FILING DATE: 2000-02-03  
NUMBER OF SEQ ID NOS: 2700  
SOFTWARE: Custom  
SEQ ID NO 1921  
LENGTH: 172  
TYPE: PRT  
ORGANISM: Homo sapiens  
PCT-US01-03800A-1921

## Alignment Scores:

Pred. No.: 1.02e-30 Length: 172  
Score: 537.00 Matches: 117  
Percent Similarity: 70.21% Conservative: 15  
Best Local Similarity: 62.23% Mismatches: 40  
Query Match: 19.66% Indels: 16  
Gaps: 3

US-09-502-945-1 (1-1552) x PCT-US01-03800A-1921 (1-172)

OY 670 GCTTCTGGGGAATGATGTCAACAAGGTGTGTGGAAGAAAGCCGATACCTGAATAA 729  
Db 1 AlaserArgGlnMetAspValThrLysValLysGlnGlnMetArgTyrGlnLeuAsnLys 20  
OY 730 ACCAACAATGAGAGAGAGAGCAAGAAAGAGCAGACAGAGTTCAGAGCAAAAACATAAC 789  
Db 21 ThrAsnMetGlnLysAspGlnAlaGlnLysGlnHisArgGlnPheArgAlaLysThrAsn 40  
OY 790 AGGATCTTGAATTAAGATCAGAGAAATAGAGAAATTGAGAAATGAGAACTGGATGAAGC 849  
|||||

Db 41 ArgAspLeuGlnLysLysAspGlnGlnLysLeuArgIleGlnLeuAspGlnSer 60  
OY 850 AAACACACTTGGAAACAGAGACAGAGAACGACCCGCGCAGAGAGAGACTGCTGAGA 909  
Db 61 LysGlnHisLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 80  
OY 910 CTAACAGAACTGTGGGCGGAATCTGAGCACAACCTGACCTCACCAGATCTGAATTAAGT 969  
Db 81 LeuThrGlnLeuLeuGlnGlnSerGlnHisGlnLeuHisLeuThrArgGlnLysAsp 100  
OY 970 CAACTCAGTCAAGAAAAAGGATATCATATGATTAATTTGGAAAGTTCAGAGAAAT 1029  
Db 101 SerIleGlnGln-----SerPheSerLysGlnAlaLysAlaGlnAlaLeuGln 116  
OY 1030 GAAGAAATGAGAGAACACTGTGCTCCAGCATGGAGAGTACATGACATCAGACAAAG 1089  
Db 117 AlaGlnGlnArgGlnGln-----GlnLeuThrGlnLys 127  
OY 1090 CTAAGGAGCTGGATTAACACAGCAGCCAGCCAGCCAGCTGCTGCTCCTCAGC 1149  
Db 128 IleGlnGlnMetGlnAlaGlnHisAspLysThrGlnAsnGlnGlnTyrLeuLeuThr 147  
OY 1150 AAGCAGAACACAGCTTCTCTGAGAGAGCAGCCCTGTGGAAGAGTGGACCGCTGCG 1209  
Db 148 SerGlnAsnThrPheLeuThrLys-----LeuLysGlnGlnCysCysThrLeuAla 164  
OY 1210 ACCCAGTTACCCAGCATCCACAA 1233  
Db 165 LysLysLeuGlnGlnIleSerGln 172

## RESULT 9

US-09-834-366-14401  
Sequence 14401, Application US/09834366  
GENERAL INFORMATION:  
APPLICANT: Bejanin, Stephane  
APPLICANT: Tanaka, Hiroaki  
APPLICANT: Dumas Milne Edwards, Jean Baptiste  
APPLICANT: Jobert, Severin  
APPLICANT: Giordano, Jean-Yves  
TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
FILE REFERENCE: 81.052.REG  
CURRENT APPLICATION NUMBER: US/09/834,366  
CURRENT FILING DATE: 2001-04-13  
PRIORITY APPLICATION NUMBER: US 60/197,873  
PRIORITY FILING DATE: 2000-04-18  
NUMBER OF SEQ ID NOS: 52153  
SOFTWARE: Patent.pm  
SEQ ID NO 14401  
LENGTH: 138  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: 126  
OTHER INFORMATION: Xaa - Asp, Glu  
US-09-834-366-14401

## Alignment Scores:

Pred. No.: 1.74e-24 Length: 138  
Score: 452.00 Matches: 96  
Percent Similarity: 74.50% Conservative: 15  
Best Local Similarity: 64.43% Mismatches: 24  
Query Match: 16.54% Indels: 14  
Gaps: 3

US-09-502-945-1 (1-1552) x US-09-834-366-14401 (1-138)

OY 709 ATGGCGTATCGCTGAATTAACCAACATGAGAGAGTGTGAGAGAAAGAGACACAGA 768  
Db 1 MetArgTyrGlnLeuAsnLysThrAsnMetGlnLysAspGlnAlaGlnLysGlnHisArg 20  
OY 769 GAGTTCAGAGCAAAACATTAACAGGATCTTGAATTAAGATCAGGAATAGAGAAATTG 828  
|||||

```
Db 21 GluPheArGAlaLysThrAsnArgSpleuGluIleLysAspIngluIleGluLysLeu 40
QY 829 AGAATGAACTGGATGAACCAACAACACTTGTGAACAGAGCAGAGCAAGCCCTG 888
    |||||
Db 41 ArgIleGluLeuAspGluSerIysGlnHisLeuGluIngluIngluIlySalAlaLeu 60
QY 889 GCCAGAGAGGATGCGCTGACACTTAACAGAACTCTGGCGAATCTGAGCACAATGCGAC 948
    |||||
Db 61 AlaArgGluGluLysLeuArgLeuThrGluLeuLeuGluIlySerGlnHisGlnLeuHis 80
QY 949 CTCACGAGATCTGAATAGTCTCAACTCAGCAGCAAAAAGATATCATGTGTAATG 1008
    |||||
Db 81 LeuThrArgIngluIlyAspSerIleGlnGln-----SerPheSerLysGlu 96
QY 1009 GGAAGCTTACAGAGAAATGAAGAAATTTGGAGAACACTGTGTCCAGCATGGAGAGTA 1068
    |||||
Db 97 AlaLysAlaGlnAlaLeuGlnAlaGlnGlnArgGluIn----- 109
QY 1069 CATGAGACGATGAACCAAGGCTTAAGCAGCTGGAT---AAGCAGACGCCAGCCACAGCC 1125
    |||||
Db 110 -----GluLeuThrGlnLysIleGlnGlnMetGluAlaGlnHisAspLysThr***Asn 127
QY 1126 CACGAGCTGTGCGACTCTCCAGCAAG 1152
Db 128 GluGlnIlyLeuLeuLeuThrSerArg 136

RESULT 10
US-60-197-873-14401
; Sequence 14401, Application US/60197873
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; APPLICANT: Dumas Malne Edwards, Jean Baptiste
; APPLICANT: Jobert, Severin
; APPLICANT: Giordano, Jean-Yves
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: 81 US1.PRO
; CURRENT APPLICATION NUMBER: US/60/197,873
; CURRENT FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 52153
; SOFTWARE: Patent.pm
; SEQ ID NO 14401
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 126
; OTHER INFORMATION: Xaa = Asp, Glu
US-60-197-873-14401

Alignment Scores:
Pred. No.: 1.74e-24 Length: 138
Score: 452.00 Matches: 96
Percent Similarity: 74.508 Conservative: 15
Best Local Similarity: 64.438 Mismatches: 24
Query Match: 16.548 Indels: 14
DB: 27 Gaps: 3

US-09-502-945-1 (1-1552) x US-60-197-873-14401 (1-138)
QY 709 ATGCGCTATCAGCTGAATAAACAACATGAGAGAGGAGGCAAGAAAAGAGCAGAGA 768
    |||||
Db 1 MetArgTyrGlnLeuAsnIlyThrAsnMetGluLysAspGluIleGluLysGlnHisArg 20
QY 769 GAGTTAGAGCAAAATACTACAGAGGATCTTGAATTAAGATCAGAAATAGGAATG 828
    |||||
Db 21 GluPheArgAlaLysThrAsnArgSpleuGluIleLysAspIngluIleGluLysLeu 40
QY 829 AGAATGAACTGGATGAACCAACAACACTTGTGAACAGAGCAGAGCAAGCCCTG 888
    |||||
Db 41 ArgIleGluLeuAspGluSerIysGlnHisLeuGluIngluIngluIlySalAlaLeu 60
```

```
QY 889 GCCAGAGAGGATGCGCTGACACTTAACAGAACTCTGGCGAATCTGAGCACAATGCGAC 948
    |||||
Db 61 AlaArgGluGluLysLeuArgLeuThrGluLeuLeuGluIlySerGlnHisGlnLeuHis 80
QY 949 CTCACGAGATCTGAATAGTCTCAACTCAGCAGCAAAAAGATATCATGTGTAATG 1008
    |||||
Db 81 LeuThrArgIngluIlyAspSerIleGlnGln-----SerPheSerLysGlu 96
QY 1009 GGAAGCTTACAGAGAAATGAAGAAATTTGGAGAACACTGTGTCCAGCATGGAGAGTA 1068
    |||||
Db 97 AlaLysAlaGlnAlaLeuGlnAlaGlnGlnArgGluIn----- 109
QY 1069 CATGAGACGATGAACCAAGGCTTAAGCAGCTGGAT---AAGCAGACGCCAGCCACAGCC 1125
    |||||
Db 110 -----GluLeuThrGlnLysIleGlnGlnMetGluAlaGlnHisAspLysThr***Asn 127
QY 1126 CACGAGCTGTGCGACTCTCCAGCAAG 1152
Db 128 GluGlnIlyLeuLeuLeuThrSerArg 136

RESULT 11
PCT-US02-09921-968
; Sequence 968, Application PC/TUS0209921
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: DAFFO, Abel
; APPLICANT: JONES, Anissa L.
; APPLICANT: TRAN, Alanna-Phung B.
; APPLICANT: DAHL, Christopher R.
; APPLICANT: GIETZEN, Darryl
; APPLICANT: CHINN, Joyce
; APPLICANT: DUFOUR, Gerard E.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: YU, Jimmy Y.
; APPLICANT: TUASON, Olivia
; APPLICANT: YAP, Pierre E.
; APPLICANT: AMSEY, Stefan R.
; APPLICANT: DAUGHERTY, Sean C.
; APPLICANT: DAM, Tam C.
; APPLICANT: LIU, Tommy F.
; APPLICANT: NGUYEN, Duy-Viet An
; APPLICANT: KLEBERD, Yael
; APPLICANT: KERSTIN JR., Edward H.
; APPLICANT: PERALTA, Careyna H.
; APPLICANT: DAVID, Marie H.
; APPLICANT: LEWIS, Samantha A.
; APPLICANT: CHEN, Alice J.
; APPLICANT: PANZER, Scott R.
; APPLICANT: HARRIS, Bernard
; APPLICANT: FLORES, Vincent
; APPLICANT: MARWAHA, Rakesh
; APPLICANT: LO, Audrey
; APPLICANT: LAN, Ruth Y.
; APPLICANT: URASHKA, Michael
; TITLE OF INVENTION: SECRETORY MOLECULES
; FILE REFERENCE: PT-1232 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09921
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/280,067; 60/280,068; 60/291,280; 60/291,849;
60/291,829; 60/299,428; 60/300,001; 60/299,776
; PRIOR FILING DATE: 2001-03-29; 2001-03-29; 2001-05-16; 2001-05-17;
2001-05-17; 2001-06-19; 2001-06-20; 2001-06-20
; NUMBER OF SEQ ID NOS: 1146
; SOFTWARE: PERL Program
; SEQ ID NO 968
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No: LT:474414.28.off1:2001MAY17
PCT-US02-09921-968
```

## Alignment Scores:

Pred. No.: 2,71e-24 Length: 94  
 Score: 449.00 Matches: 81  
 Percent Similarity: 93.10% Conservative: 0  
 Best Local Similarity: 93.10% Mismatches: 0  
 Query Match: 16.30% Indels: 6  
 DB: 1 Gaps: 1

US-09-502-945-1 (1-1552) x PCT-US02-09921-968 (1-94)

OY 1223 CTGGGTAAC-----TGGCGCCAGCCGGTCCACCTTTCGACAGG 1182  
 DB 8 LeuGIaAnIIleValArgProHisLeuTyrValArgSerArgSerThrSerSerAspArg 27  
 OY 1181 CTCTGCCCTCCAGAGAAAGCGCTTCCTGCTGAGAGAGTGCAGCAGCTCGGGCT 1122  
 DB 28 LeuCySLeuSerArgArgSerTyrPheCySLeuLeuArgSerCySthSerCyStrpAla 47  
 OY 1121 GTGGCCCTGGCTGTCTTCACAGCTGCCTTAGCCTTTCCTTCATGCTCATGTAAGTCTC 1062  
 DB 48 ValAlaIrrpleuCySLeuSerSerCySLeuSerLeuCySphleuLeuValSerCyStrhIeu 67  
 OY 1061 CCATGCTGACACACTGTCTCCATTCATTCATTTCTCTGTGAATTCCTCCAAATTTA 1002  
 DB 68 ProCyStrPrrHhScySerSerSerSerSerPheLeuLeuCySAsnPhroAsnleu 87  
 OY 1001 TCATATGTATACCTTTTCT 981  
 DB 88 SerTyValTyLeuPheSer 94

## RESULT 12

US-09-834-366-21271  
 : Sequence 21271, Application US/09834366  
 : GENERAL INFORMATION:  
 : APPLICANT: Benjamin, Stephane  
 : APPLICANT: Tanaka, Hiroaki  
 : APPLICANT: Dumas Milne Edwards, Jean Baptiste  
 : APPLICANT: Jobert, Severin  
 : APPLICANT: Giordano, Jean-Yves  
 : TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
 : FILE REFERENCE: 81.US2.REG  
 : CURRENT APPLICATION NUMBER: US/09/834,366  
 : CURRENT FILING DATE: 2001-04-13  
 : PRIOR APPLICATION NUMBER: US 60/197,873  
 : PRIOR FILING DATE: 2000-04-18  
 : NUMBER OF SEQ ID NOS: 52153  
 : SOFTWARE: Patent.pm  
 : SEQ ID NO 21271  
 : LENGTH: 61  
 : TYPE: PRT  
 : ORGANISM: Homo sapiens  
 US-09-834-366-21271

## Alignment Scores:

Pred. No.: 4.16e-14 Length: 61  
 Score: 310.00 Matches: 61  
 Percent Similarity: 98.39% Conservative: 0  
 Best Local Similarity: 98.39% Mismatches: 0  
 Query Match: 11.35% Indels: 1  
 DB: 22 Gaps: 0

US-09-502-945-1 (1-1552) x US-09-834-366-21271 (1-61)

OY 1078 ATGAAGCAAGGCTAAGCAGCTGATTAAGCACAGCCAGCCAGCCAGCAGCTGCTG 1137  
 DB 1 MetLysGlnAArgLeuArgGlnLeuAspLysHisSerGlnAlaThrAlaGlnGlnLeuVal 20  
 OY 1138 CAGCTCTCAGACAGACAGACAGCTTCTCTGAGAGAGCAGAGCTGTGGAAAGAGTG 1197  
 DB 21 GlnLeuLeuSerLysGlnAsnGlnLeuLeuLeuGlnArgInser-CysArgLysArgTr 40  
 OY 1198 GACCGGCTGCGAGCCAGCTAACCCAGCATGCCACAATGTGATGCTGACCTGATGGAAC 1257  
 DB 1198 GACCGGCTGCGAGCCAGCTAACCCAGCATGCCACAATGTGATGCTGACCTGATGGAAC 1257

DB 40 pThrGlyCySgLyProSerTyProAlaCySHisAsnLeuIleAlaAspLeuAspGlyTh 60

OY 1258 AGAG 1261

DB 60 rGlu 61

## RESULT 13

US-60-197-873-21271

: Sequence 21271, Application US/60197873

: GENERAL INFORMATION:

: APPLICANT: Benjamin, Stephane

: APPLICANT: Tanaka, Hiroaki

: APPLICANT: Dumas Milne Edwards, Jean Baptiste

: APPLICANT: Jobert, Severin

: APPLICANT: Giordano, Jean-Yves

: TITLE OF INVENTION: ESTs and Encoded Human Proteins.

: FILE REFERENCE: 81.US1.PRO

: CURRENT APPLICATION NUMBER: US/60/197,873

: CURRENT FILING DATE: 2000-04-18

: NUMBER OF SEQ ID NOS: 52153

: SOFTWARE: Patent.pm

: SEQ ID NO 21271

: LENGTH: 61

: TYPE: PRT

: ORGANISM: Homo sapiens

US-60-197-873-21271

## Alignment Scores:

Pred. No.: 4.16e-14 Length: 61  
 Score: 310.00 Matches: 61  
 Percent Similarity: 98.39% Conservative: 0  
 Best Local Similarity: 98.39% Mismatches: 0  
 Query Match: 11.35% Indels: 1  
 DB: 27 Gaps: 0

US-09-502-945-1 (1-1552) x US-60-197-873-21271 (1-61)

OY 1078 ATGAAGCAAGGCTAAGCAGCTGATTAAGCACAGCCAGCCAGCCAGCAGCTGCTG 1137  
 DB 1 MetLysGlnAArgLeuArgGlnLeuAspLysHisSerGlnAlaThrAlaGlnGlnLeuVal 20  
 OY 1138 CAGCTCTCAGACAGACAGACAGCTTCTCTGAGAGAGCAGAGCTGTGGAAAGAGTG 1197  
 DB 21 GlnLeuLeuSerLysGlnAsnGlnLeuLeuLeuGlnArgInser-CysArgLysArgTr 40  
 OY 1198 GACCGGCTGCGAGCCAGCTAACCCAGCATGCCACAATGTGATGCTGACCTGATGGAAC 1257  
 DB 40 pThrGlyCySgLyProSerTyProAlaCySHisAsnLeuIleAlaAspLeuAspGlyTh 60  
 OY 1258 AGAG 1261  
 DB 60 rGlu 61

## RESULT 14

US-09-621-976-5141  
 : Sequence 5141, Application US/09621976  
 : GENERAL INFORMATION:  
 : APPLICANT: Dumas Milne Edwards, J.B.  
 : APPLICANT: Jobert, S.  
 : APPLICANT: Giordano, J.Y.  
 : TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
 : FILE REFERENCE: GENSET.054PR2  
 : CURRENT APPLICATION NUMBER: US/09/621,976  
 : CURRENT FILING DATE: 2000-07-21  
 : NUMBER OF SEQ ID NOS: 19335  
 : SOFTWARE: Patent.pm  
 : SEQ ID NO 5141  
 : LENGTH: 62  
 : TYPE: PRT  
 : ORGANISM: Homo sapiens  
 US-09-621-976-5141

## Alignment Scores:

Pred. No.: 2.26e-13 Length: 62  
Score: 300.00 Matches: 59  
Percent Similarity: 98.33% Conservative: 0  
Best Local Similarity: 98.33% Mismatches: 0  
Query Match: 10.98% Indels: 1  
DB: 20 Gaps: 0

US-09-502-945-1 (1-1552) x US-09-621-976-5141 (1-62)

OY 1078 ATGAGCAAAAGGCTAAGCAGCTGATTAAGCAGCCAGCCAGCCAGCAGCTGGTG 1137  
|||||  
Db 1 MetLysGlnArgLeuArgGlnLeuAspLysHisSerGlnAlaThrAlaGlnGlnLeuVal 20

OY 1138 CAGCTCCTCAGACAGAGAACACAGCTTCTCCTGAGAGCCAGCCTGTGGAAGAGTG 1197  
|||||  
Db 21 GlnLeuLeuSerLysGlnAsnGlnLeuLeuGlnArgGlnSer-CysArgLysArgTr 40

OY 1198 GACCGGCTGCGGACCCAGTTTACCAGCATGCCCAATCTGATTGCTGACCTGGATGGA 1255  
|||||  
Db 40 pThrGlyCysGlyProSerTyrProAlaCysHisAsnLeuIleAlaAspLeuAspGly 59

#### RESULT 15

US-60-147-499-5141

; Sequence 5141, Application US/60147499

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Jobert, S.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: ESTs and Encoded Human Proteins.

; FILE REFERENCE: GENSET.054PR2

; CURRENT APPLICATION NUMBER: US/60/147,499

; CURRENT FILING DATE: 1999-08-05

; NUMBER OF SEQ ID NOS: 19335

; SOFTWARE: Patent.pm

; SEQ ID NO 5141

; LENGTH: 62

; TYPE: PRT

; ORGANISM: Homo sapiens

US-60-147-499-5141

#### Alignment Scores:

Pred. No.:	2.26e-13	Length:	62
Score:	300.00	Matches:	59
Percent Similarity:	98.33%	Conservative:	0
Best Local Similarity:	98.33%	Mismatches:	0
Query Match:	10.98%	Indels:	1
DB:	27	Gaps:	0

US-09-502-945-1 (1-1552) x US-60-147-499-5141 (1-62)

OY 1078 ATGAGCAAAAGGCTAAGCAGCTGATTAAGCAGCCAGCCAGCCAGCAGCTGGTG 1137  
|||||  
Db 1 MetLysGlnArgLeuArgGlnLeuAspLysHisSerGlnAlaThrAlaGlnGlnLeuVal 20

OY 1138 CAGCTCCTCAGACAGAGAACACAGCTTCTCCTGAGAGCCAGCCTGTGGAAGAGTG 1197  
|||||  
Db 21 GlnLeuLeuSerLysGlnAsnGlnLeuLeuGlnArgGlnSer-CysArgLysArgTr 40

OY 1198 GACCGGCTGCGGACCCAGTTTACCAGCATGCCCAATCTGATTGCTGACCTGGATGGA 1255  
|||||  
Db 40 pThrGlyCysGlyProSerTyrProAlaCysHisAsnLeuIleAlaAspLeuAspGly 59

Search completed: March 21, 2003, 13:42:51  
Job time : 174.803 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - protein search, using frame\_plus.n2p model

Run on: March 21, 2003, 13:00:02 ; Search time 16.6182 Seconds  
(without alignments)  
9985.861 Million cell updates/sec

Title: US-09-502-945-1  
Perfect score: 2732  
Sequence: 1 cttctgcatgcatccgagaa.....aaatgacttttaagaagaa 1552

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 221153 seqs, 53462247 residues  
Total number of hits satisfying chosen parameters: 442306

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+n2p.model -DEV=x1p  
-O/cgn2\_1/USPTO.spool/US09502945/runal\_14032003\_101101\_19301/app\_query.fasta\_1.10979  
-DB=published.Applications\_AA -QEXT=fastan -SUFFIX=rapb -MINMATCH=0.1  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62  
-TRANS=human40.csl -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09502945.0CGN.1.1.60.0runal\_14032003\_101101\_19301  
-NCPU=6 -ICPU=3 -NO\_XMAP -LARGEQUERY -NEG\_SCORES=0 -WRITE -LONGLOG  
-DEV\_TIMEDOUT=120 -WARN\_TIMEDOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELOP=6 -DELEXT=7

## Database : Published Applications AA:\*

1: /cgn2\_6/prodata/2/pubppaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/prodata/2/pubppaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/prodata/2/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/prodata/2/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/prodata/2/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/prodata/2/pubppaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/prodata/2/pubppaa/PCTUS\_PUBCOMB.pep:\*  
8: /cgn2\_6/prodata/2/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/prodata/2/pubppaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/prodata/2/pubppaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/prodata/2/pubppaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/prodata/2/pubppaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/prodata/2/pubppaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/prodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	242.5	8.9	2139	9	US-10-023-219-4
2	242.5	8.9	2139	9	US-09-727-384-6
3	238.5	8.7	868	9	US-09-884-001-19
4	231.5	8.5	2125	10	US-09-919-172-29

5	229.5	8.4	909	9	US-09-925-299-988	Sequence 988, App
6	229.5	8.4	909	10	US-09-925-299-988	Sequence 988, App
7	229	8.4	2354	9	US-09-820-843A-113	Sequence 113, App
8	228	8.3	2310	9	US-09-991-496-120	Sequence 120, App
9	228	8.3	2310	9	US-09-820-843A-114	Sequence 114, App
10	228	8.3	2310	10	US-09-874-923-120	Sequence 120, App
11	226.5	8.3	888	9	US-09-893-519A-73	Sequence 73, App
12	224.5	8.2	1805	9	US-09-820-843A-73	Sequence 73, App
13	222.5	8.1	1045	10	US-09-815-242-10617	Sequence 10617, A
14	220	8.1	2053	9	US-10-017-216-2	Sequence 2, App
15	218.5	8.0	660	10	US-09-864-761-47959	Sequence 47959, A
16	218.5	8.0	1945	9	US-09-927-597-2	Sequence 2, App
17	218.5	8.0	1597	9	US-09-927-597-4	Sequence 4, App
18	215.5	7.9	1597	9	US-10-017-216-6	Sequence 6, App
19	215.5	7.9	2055	9	US-10-017-216-4	Sequence 4, App
20	215	7.9	1641	9	US-10-017-216-5	Sequence 5, App
21	214.5	7.9	1958	12	US-10-028-946-4	Sequence 4, App
22	214.5	7.9	2054	12	US-10-028-946-2	Sequence 2, App
23	213	7.8	900	12	US-10-071-751-21	Sequence 21, App
24	208	7.6	443	1	US-08-325-278-6	Sequence 6, App
25	207.5	7.6	691	9	US-10-028-072-16	Sequence 16, App
26	207.5	7.6	691	9	US-10-121-049-16	Sequence 16, App
27	207.5	7.6	691	9	US-10-123-904-16	Sequence 16, App
28	207.5	7.6	691	9	US-10-140-470-16	Sequence 16, App
29	207.5	7.6	691	9	US-10-175-746-16	Sequence 16, App
30	207.5	7.6	691	9	US-10-176-918-16	Sequence 16, App
31	207.5	7.6	691	9	US-10-176-921-16	Sequence 16, App
32	207.5	7.6	691	9	US-10-137-865-16	Sequence 16, App
33	207.5	7.6	691	9	US-10-140-474-16	Sequence 16, App
34	207.5	7.6	691	9	US-10-142-431-16	Sequence 16, App
35	207.5	7.6	691	9	US-10-143-114-16	Sequence 16, App
36	207.5	7.6	691	9	US-10-140-002-16	Sequence 16, App
37	207.5	7.6	691	9	US-10-142-419-16	Sequence 16, App
38	207.5	7.6	691	9	US-10-123-262-16	Sequence 16, App
39	207.5	7.6	691	9	US-10-142-423-16	Sequence 16, App
40	207	7.6	830	9	US-10-033-245-7	Sequence 7, App
41	207	7.6	830	9	US-10-033-223-7	Sequence 7, App
42	207	7.6	830	9	US-10-033-167-7	Sequence 7, App
43	207	7.6	830	9	US-10-033-244-7	Sequence 7, App
44	207	7.6	830	9	US-10-033-435-7	Sequence 7, App
45	207	7.6	830	9	US-10-032-990-7	Sequence 7, App

## ALIGNMENTS

RESULT 1  
US-10-023-219-4  
; Sequence 4, Application US/10023219  
; Publication No. US20030032592A1  
; GENERAL INFORMATION:  
; APPLICANT: Myriad Genetics, Inc.  
; APPLICANT: Cimbora, Daniel M.  
; APPLICANT: Helichman, Karen  
; APPLICANT: Bartel, Paul L.  
; TITLE OF INVENTION: Protein-Protein Interactions  
; FILE REFERENCE: 2318-282-II  
; CURRENT APPLICATION NUMBER: US/10/023, 219  
; CURRENT FILING DATE: 2001-12-20  
; PRIOR APPLICATION NUMBER: US 60/256, 983  
; PRIOR FILING DATE: 2000-12-21  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 2139  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-023-219-4

Alignment Scores:  
Pred. No.: 3 21e-10  
Score: 242.50  
Percent Similarity: 41.86%  
Best Local Similarity: 24.21%  
Length: 2139  
Matches: 107  
Conservative: 78  
Mismatch: 158

Query Match:	8.88%	Indels:	99
DB:	9	Gaps:	18
US-09-502-945-1 (1-1552) x US-10-023-219-4 (1-2139)			

[illegible][illegible]

## RESULT 2

; Sequence 6, Application US/09727384

Patent No. US20020098511A1

APPLICANT: Myriad Genetics, Inc.

APPLICANT: Cimborra, Daniel M.

APPLICANT: Mauck, Kimberly

TITLE OF INVENTION: Protein-protein interactions

FILE REFERENCE: 2318-2/1

; CURRENT FILING DATE: 2000-12-01

PRIOR FILING DATE: 1999-12-02

PRIOR FILING DATE: 1999-12-02

PRIOR APPLICATION NUMBER: US 60/185,056

; NUMBER OF SEQ ID NOS: 8

; SEO ID NO 6

TYPE: PRT

ORGANISM: Homo sapiens

Pred. No.: 3.

Percent Similarity: 41

Best LOCAL SIMILARITY: 24

DB: 10

US-09-502-945-1 (1-1552) x US-09-727-384-6 (1-2139)

10 GCATTCGAGAGCTAAACCTTACTTATTGAGGAAGAAAGTGTGAATTCAGGAATCCCAATTG 69

[illegible][illegible]

070 ZELANDIALENTAULCEUINOTYOLHATAVAALPEUZYOTA

130 CTAAGCATTAAGAATTCTTCTGGCTGCTAATACITGTAAACCGTGTGGTGGTCTTTGT 18



Db 694 ----- 694  
QY 190 TTGAATGCTGCACATGAAGCTGTTTCCCAACCCACTAATGTTCAATGCGAG 249  
Db 695 -----AlaHisHisGluAlaThrCysArgHisGluGluGluLysGluLeuGln 711  
QY 250 ACCATCGAAGAGCTGTTAAAGAAAGATGACTGATGCTGCACATA----- 297  
Db 712 Val-----LysLeuGluGluGluLysThrHisLeuGluGluLysLeuArgLeuHis 729  
QY 298 ---GTTTCGTAAGAGAGCGCTTGCGACATACGACCAAGAGCAAGCTTATGAA 354  
Db 730 GluMetGluLeuLysAlaArgLeuThrGlnAlaGln-----AlaSerPheGlu 745  
QY 355 CAGGTGAACAAAGTTTCCAAATATCTGAGAGGCCAATTTGAAAAAACCAAGCTTTA 414  
Db 746 ArgGluArgGluGluLeuGlnSerSerAlaTrpThrGlu---GluLysValArgGluLeu 764  
QY 415 ATCCAGTGTGACCGATTGAGAGAGCGTGGAGAGCGGCGGCGACTTGAAGAAAGAA 474  
Db 765 ThrGln-----GluLeuGlu-----GlnPheHisGluGluGln 775  
QY 475 CTTCGATCTCAGACAGAGAAAAGGCCATTGAGAAAAGCATGATGAAAAGAAATAACG 534  
Db 776 LeuThrSerLeuValGluLysHisThrLeuGlnLysGluGluLeuArgLysGluLeu 795  
QY 535 AAAGAA-----AGGAGTACATGGGATCAAAAGATGTTGATC 570  
Db 796 GluLysHisGlnArgGluLeuGlnGluGluLysArgGluLysMetGluThrGlu----- 812  
QY 571 TTGCTCAGAAATATGTCACCTGACGAGCCAG-----GTGAAAAAGTTACA 618  
Db 813 CysAsnArgArgThrSerGlnLeuAlaGlnPheGlnSerArgPcyGlnLysValThr 832  
QY 619 AAGGAAAAGATTCACTATTATCACTGAGAGAAATCAAGCCAGCTGCTTCGCG 678  
Db 833 GluArgCysGluSerAlaLeuGlnSerLeuGluGluArgLysArgGln----- 848  
QY 679 GAAATGAGTGCACAAAGGTGTGGAGAAATGCGCTTCAGCTGATAAACAACACATG 738  
Db 849 -----GluLeuLysAspLeuGlnGluGlnArgGluGluLysSerGlnTrpGluPhe 866  
QY 739 GAGAGAGATGAGCAGAAAAGAGCAGACAGAGTTTCAGAGCAAAAACATAAGAGGATCTT 798  
Db 867 GluLysAspGluLeuThrGlnGluCysAlaGluAlaGlnGluLeuLysGluThrLeu 886  
QY 799 GAA-----ATTAAAGATCAGAAATAGAGAAATTAGAAATAGAA 837  
Db 887 LysArgGluLysThrThrSerLeuValLeuThrGlnGluArgGluMetLeuGln----- 904  
QY 838 CTGATGTAACCAACTGTGGACACAGAGCGAGAGGCCCTGGCCAGAGAG 897  
Db 905 ---LysThrTyrLysGluHisLeu-----AsnSerMetValValGluArgGln 919  
QY 898 GAGTGCCCTGAGAA---CTAACAGAACTGCTGGCGAATCTGACACCACTCACCTCACCC 954  
Db 920 GlnLeuLeuGlnAspLeuGlnAspLeuArgAsnValSerGluThrGlnGlnSerLeuLeu 939  
QY 955 AGATCTGAATAGCTCAACTAGTCAGAGAAAAGGTTACATATGATTAATTTGGGAAAG 1014  
Db 940 SerAspGlnIleLeuGlnLeuLysSerSerHisLys----- 951  
QY 1015 TTACAGAGAAAGAAATGGAATTTGGAGAGACAGTGTCTCCAGCATGGAGAGTACATAG 1074  
Db 952 -----ArgGluLeuArgGluArgGluValLeuCysGlnAlaGlyAlaSerGluGln 969  
QY 1075 ACAGTGAAGCAAGAGCTAGAGCTGGAT---AAGCAGACAGAGCCAGACGCCAGAG 1131  
Db 970 LeuAlaSerGlnArgLeuGlnArgLeuGlnMetGlnHisAspGlnGluArgGlnMet 989  
QY 1132 CTGTGTGAGCTCTCAGACAGCAGAAC-----CACCTTCTCTCTG 1170  
Db 990 MetSerLysLeuLeuAlaMetGlnAsnIleHisLysAlaThrCysGluThrAlaAspArg 1009

QY 1171 GAGAGCAGAGCTGTCTCGAAGAGGTGAGACCGGCTGGCGACCCAGTATCCAGCATGCA 1230  
Db 1010 GluArgAlaGluMetSerThrGluLeuSerArgLeuGlnSerLysIleLysGluMetGln 1029  
QY 1231 CAATCT 1236  
Db 1030 GlnAla 1031  
RESULT 3  
US-09-884-001-19  
: Sequence 19, Application US/09884001  
: Publication No. US20020182656A1  
: GENERAL INFORMATION:  
: APPLICANT: Bird, Timothy A.  
: APPLICANT: Peschon, Jacques J.  
: APPLICANT: Sims, John E.  
: APPLICANT: Virca, G. Duke  
: APPLICANT: Willis, Cynthia R.  
: TITLE OF INVENTION: Methods for Regulating Vascularization Using GEF  
: TITLE OF INVENTION: Containing NEK-Like Kinase (GNK)  
: FILE REFERENCE: Immunex GNK/SGNK PCT  
: CURRENT FILING DATE: 2001-06-18  
: PRIOR APPLICATION NUMBER: US/09/884,001  
: PRIOR FILING DATE: 1998-12-18  
: NUMBER OF SEQ ID NOS: 19  
: SOFTWARE: Patencin Ver. 2.0  
: SEQ ID NO 19  
: LENGTH: 868  
: TYPE: PRT  
: ORGANISM: Homo sapiens  
US-09-884-001-19  
Alignment Scores:  
Pred. No.: 5,34e-10 Length: 868  
Score: 238.50 Matches: 123  
Percent Similarity: 39.08% Conservative: 99  
Best Local Similarity: 21.65% Mismatches: 198  
Query Match: 8.73% Indels: 148  
Gaps: 18  
US-09-502-945-1 (1-1552) x US-09-884-001-19 (1-868)  
QY 37 GAGGAAAAGTGTGAATGAGAGATCCCAATTGAAGTTTGGAGAGCACTTACTGAA 96  
Db 27 GluGluLysGluGluLys-----LysLeuSerGluSerArgHisGlnGlnAla 45  
QY 97 TATCAGAACTGTGAGAGATCTTAAAGAGCAACTAAGCATTAAGAAATTTCTTGCTG 156  
Db 46 AlaThrThrGlnLeuGlnGlnLeuHisGlnLysAlaLysArgGlnGlnValLeuAla 65  
QY 157 GCTAATACTTAAACCGTGTGGTGGCTTTGTTGAATGCTCAGCATGAAGCTGTT 216  
Db 66 -----ArgAlaValGlnGluAla 73  
QY 217 CTT-----TCCCAACCCACTAATGTTTCATATGCAGACCATCGAAGAA----- 261  
Db 74 LeuValArgGluLysAlaAlaLeuGlnValArgLeuGlnAlaValGluArgAspArgGln 93  
QY 262 ---CTGTTAAAGAAAGATGACTGTGCTGCACTAGTTTCCGTAAAGAGCAGCTTG 318  
Db 94 AspLeuAlaGlnGlnLeuGlnGlyLeuSerSerAlaLysGluLeuLeuGlnSerLeu 113  
QY 319 GCAGATACGACAGCAAGA-----GAA 339  
Db 114 PheGluAlaGlnGlnGlnAsnSerValIleGluValThrLysGlnLeuGlnValGln 133  
QY 340 GCAAGTCTTATGAACAGAGTAAACAAGTTTTCGAA-----ATATCTGAG 384  
Db 134 IleGlnThrValThrGlnAlaLysGlnValIleGlnGlnGluValArgCysLeuLysLeu 153  
QY 385 GAAGCAATTTTGAAGAAAACCAAGCT-----TTA 414

```

Db 154 GluLeuAspThrGlnArgSerGlnAlaGlnGlnIuArgAspAlaAlaArgGlnLeu 173
OY 415 ATCCAGTGTGACCACTTGAGAGAG---GACCTGGAGAGCGAGCGGACCACTTGAAAA 471
Db 174 AlaGlnAlaGlnIuArgSerGlnAlaLeuGlnIuGlnLysAlaAlaAlaGlnLys 193
OY 472 GAACCTGCACTCAGCAAGAGAAAGGCCATTCAGCAAAACATGATGAAAGAAATA 531
Db 194 GluValAsnGlnLeuArgGlnIuArgGlnLysGlnIuArgSerGlnIuGlnIuLeu 213
OY 532 AGCAAA-----GAAAGGAGTACATGGCATCAAGATGTTGATCTTGCT 576
Db 214 AlaLysAlaLeuGlnIuArgSerGlnIuArgGlnLysMetGlnIuLeuIuArgLysGln 233
OY 577 CAGAAATAT----- 585
Db 234 GlnGlnIuThrGlnMetGlnAlaIleGlnAlaGlnIuArgGlnIuArgGlnAlaGln 253
OY 586 -----GCCCACTGAGGCCCGGAGTGAAGGTTACA----- 618
Db 254 SerAlaLeuGlnIuMetGlnIuGlnIuThrGlnLysGlnIuArgValSerLeuLeuGlnIuThr 273
OY 619 -----AAGCAAAAGATTTGAGCTATTATCACTGAGCAAAATTCAAAGC 663
Db 274 LeuLeuGlnIuThrGlnLysGlnIuLeuAlaAspAlaSerGlnIuGlnIuArgGln 293
OY 664 CAGCTGGCTTCGGGAAATGATGTCACAAAGTGTGTGAGAAATGCGCTATCAGCTG 723
Db 294 AspMetLysValGlnLysLeuLysGlnIuGlnIuThrGlnIuLeuGlnIuThrGlnIuLeu 313
OY 724 AATTAACAACATGAGAGAGATGAGCGCAAAAGAGCAGACAGATTCAAGCA--- 780
Db 314 GlnGlnIuAlaGlnIuArgGlnIuLeuLysGlnIuAlaAlaArgGlnIuIleArgSerLeuAlaAla 333
OY 781 -----AAACTAACAGGATCTTGAATTAAGATCAG 813
Db 334 LeuGlnIuGlnIuSerSerLeuLeuGlnIuAspLysMetSerLeuGlnIuLysGlnIuValGln 353
OY 814 GAATAGAGAAATTTGAAATAGAACTGATGAAGCAAAACACTTGGAAAGAGCAG 873
Db 354 AspLeuLysSerGlnIuLeuValAlaGlnAspSerGlnIuArgLeuValGlnIuGlnIuVal 373
OY 874 CAGAGGCAAGCCCTGGCCAGAGAGAGTGCCTGAGACTTAACAGAACTGCTGGCGAATCT 933
Db 374 GlnGlnIuLysLeuArgGlnIuThrGlnIuIuTyAsnArgIleGlnIuLysGlnIuLeuArgGln 393
OY 934 GAGCACAACCTGCACCTCCAGATCTGAATAGCTCAA-----CTCAGTCA 981
Db 394 LysAlaSerLeuThrLeuSerLeuMetGlnLysGlnIuArgLeuValLeuGlnIuGln 413
OY 982 GAAAAAAGTATACATATGATAATTGGAAAGTTCACAGAGAAATGAAGAATTGGAG 1041
Db 414 AlaAspSerIleArgGlnIuGlnIuLeuSerAlaLeuArgGlnAspMetGlnIuAlaGln 433
OY 1042 GAACATGTGTCCAGCATGGAGAGATACATGACACATGACAGCAAAAGCTTAAGCGACTG 1101
Db 434 GlnGlnIuLysGlnIuLeuSerAlaGlnIuMetGlnIuLeuLysGlnIuValLysGlnIuLys 453
OY 1102 GAT-----AAGCAGCAGCCAGGCC 1119
Db 454 GluAlaLysPheLeuAlaGlnIuAlaGlnIuLeuGlnIuLeuGlnIuLysAlaSerHisIle 473
OY 1120 ACAGCCAGCAGCTGTGGAGTCTC-----AGCAAGAGAAACCACTTCTC 1167
Db 474 ThrGlnGlnIuLeuArgAlaSerLeuTrpAlaGlnIuAlaLysAlaAlaGlnIuGln 493
OY 1168 CTGAGAGCAGCAGCTGTGGAGAGAGTGGAGCGGCTG----- 1206
Db 494 LeuArgLeuArgSerThrGlnSerGlnIuGlnIuAlaLeuAlaGlnIuGlnIuProGln 513
OY 1207 -----CGAGCCAGTTACCCAGCATGCCCAATCTGATTCCTCAGCTGGA 1251

```

```

Db 514 AsnGlnAlaGlnIuAlaGlnIuLeuAlaSerLeu-TyrSerAlaLeu----- 539
OY 1252 TGGACAGAGTGAATTAATGAATTTACAAAGAGATATTTCATTCACTTCGTTAGACT 1311
Db 530 -----GlnGlnIuAlaLeuGlnLys 535
OY 1312 AATATGCCACAGCAGCAGCAGCTTCC---CAGGATGACAGCGGCTCAGCTGAGTGG 1368
Db 535 ValCysGlnSerArgProIuLeuSerGlnIuGlnIuAspSerAlaIleProSerValTrpGln 555
OY 1369 GCGTGTCTCTATCAACGCGGG 1390
Db 555 YLeuGlnProAspGlnAsnGln 562

RESULT 4
US-09-919-172-29
; Sequence 29, Application US/09919172
; Patent No. US20020119463A1
; GENERAL INFORMATION:
; APPLICANT: Fafis, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 29
; LENGTH: 2125
; TYPE: PRF
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc. feature
; OTHER INFORMATION: Incyte ID No. US20020119463A1 3774181CD1
US-09-919-172-29

Alignment Scores:
Pred. No.: 2,13e-09 Length: 2125
Score: 231.50 Matches: 108
Percent Similarity: 41.20% Conservative: 98
Best Local Similarity: 21.60% Mismatches: 153
Query Match: 8.47% Indels: 141
DB: Gaps: 20

US-09-502-945-1 (1-1552) x US-09-919-172-29 (1-2125)
OY 37 GAGGAAAGGTGTAATTTGAGGAATCCCAATTGAAGTT-----TTGAGGAACGAC 87
Db 628 LysGlnLysSerArgValGlnGlnIuLeuProLysValArgGlnAlaAlaGlnIuAsnGln 647
OY 88 TTAGCTGAATATAGAGAACTTGTGAAGATCTTAAGAGCA----- 129
Db 648 LeuArgLysGlnIuArgAsnValGlnAspIleSerLeuGlnIuLysIleArgAlaGlnIuSer 667
OY 130 -----CTAAGCATTAAGAAATTTCTTCTG 153
Db 668 GluAlaLysGlnIuTyArgArgGlnIuLeuGlnIuThrIleValArgGlnIuLysGln 685
OY 154 GCTGCTAATTAATCTTAACCGCTGTGGTCTTGTGTAATGTGCTGACATGAAGCT 213
Db 686 AlaGlnIuArgGlnIuLeuGlnIuArgValArgGlnIuLeuThrIleGlnIuAlaGlnIuLysArgAla 705
OY 214 GTTCTTCCCAAAACCATCACTAATGTTCAATATCAG----- 249
Db 706 AlaValGlnIuLysLeuLeuAsnPheArgAsnGlnIuLeuGlnIuAsnThrPheThrArg 725
OY 250 ---ACCATGAAAGACTGTTAAAGAAAGAGATGACTTGATGTCTGACATGTTCCGTA 306
Db 726 ArgThrLeuGlnIuAspHisLeuLysArgLysAsp----- 736
OY 307 AGGAGCAGCTTGGCAGATGCGCAGCAAGAGAAAGCAAGTGTCTTATGACAGGTGAACAA 366

```

```

Db      1074  GluSnIleValLeuGluYlGlnInThrIleGlnIArgCysGluAlaLeuYsIleGln 1093
RESULT 5
US-09-925-299-988
: Sequence 988, Application US/09925299
: Publication No. US20030040617A9
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
: FILE REFERENCE: PA102
: CURRENT APPLICATION NUMBER: US/09/925,299
: PRIORITY FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: PCT/US00/05883
: PRIOR FILING DATE: 2000-03-08
: PRIOR APPLICATION NUMBER: 60/124,270
: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 1556
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 988
: LENGTH: 909
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (32)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (41)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (47)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (48)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (52)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (58)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (62)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (125)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (632)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (851)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-988
: Alignment Scores:
: Pred. No.:
: Score: 2,53e-09 Length: 909
: Percent Similarity: 41.63% Matches: 100
: Best Local Similarity: 22.03% Conservative: 89
: Query Match: 8.40% Mismatches: 196
: DB: 9 Indels: 69
: Gaps: 13
US-09-502-945-1 (1-1552) x US-09-925-299-988 (1-909)
OY      4  CTGGATGCATCCGAGAACTTAATAACTTACTTATGAGGAAAGTGGAATTTAGGAAATCC 63
Db      354  LeuGluGluYsGluYsLeuAlaThrGluGlnGluYsAlaValAlaYsSer 373
OY      64  CAATGAAAGTTTGTGAGGAAACGACTGAATTCAGAGAACTTGT----- 111
Db      374  LysLeuArgGluLeuYsAlaYsGluMetAlaAlaGluYsAlaYsAlaAlaAlaGlyGlu 393

```





```

OY 301 TCCGTAGAGCAGCTTGGCAGATACGACAAAGAACAGTCTTATGACAGCTG 360
    ::::: ::::: ::::: :::::
DB 960 ArgLeuAlaAlaAspArgSpGluAlaArgGlnGlnLeuAlaAlaAsnAlaGluGluLeu 979
OY 361 AACAAGCTTTGCAATATCTGAGAACCAATTGGAAAAAACAAGCTTTATTCAC 420
    ::::: ||| ::::: ::::: |||
DB 980 GlnGlnArgLeuAspThrAlaThrGlnGlnArgAlaGlnLeuAlaGlnLeuAlaArg 999
OY 421 TGT-----GACAGTTGAGAAAGAGCTGAGAGCGAGCGAGCTTGA 468
    ||||| ::::: |||||
DB 1000 LeuAlaAlaAspArgSpGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGluGln 1019
OY 469 AAA-----GAACTGCACTCGACCAAGAAAGG 498
    ::::: |||||
DB 1020 GlnArgLeuAspThrAlaThrGlnGlnArgAlaGlnLeuGlnAlaValAlaArgLeu 1039
OY 499 GCCATTGAGAAACATGATGAAAAAGAAATACGAAGAAAGGACGTCATGGCATCA 558
    ||| ::::: |||
DB 1040 AlaAlaAspArgSpGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGluGlnGln 1059
OY 559 AAGATGTTATCTTGTCTGAGATATTGCCAAGTGGAGCGGAGTGGAAAGTTACA 618
    ::::: ::::: |||||
DB 1060 ArgLeuAspThrAlaThrGlnGlnArgAlaGlnLeuGlnAlaArgValAlaArgLeuAla 1079
OY 619 AAGGAAAGATTTTCAGCTATTATCAACTG-----GAGAAATTCAAAGCCAG 666
    ::::: |||
DB 1080 AlaAspArgSpGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGluGlnGlnArg 1099
OY 667 CAG-----GCTTCGGGAAATGATGTCACAAAGGTGTGCA 705
    ||| ||| |||
DB 1100 LeuAspThrAlaThrGlnGlnArgAlaGlnLeuGlnAlaGlnValAlaArgLeuAlaAla 1119
OY 706 -----GAAATGCGCTATCAGCTG-----AATAAACCAACTGGAAG----- 744
    ||| ||| |||||
DB 1120 AspGlyAspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGluGlnGlnArgLeu 1139
OY 745 GATGAGCGCAAGAAAGACACAGAGCTTCAGACAAAACTACAGCGATCTTGAAT 804
    ||| ||| :::::
DB 1140 AspThrAlaThrGlnGlnArgAlaGlnLeuGlnAlaArgValAlaArgLeuAlaAlaAsp 1159
OY 805 AAGAGCAAGAAATAGCAAAATAGCAATAGCAATGAGATGAGCAACCAACTGGAA 864
    ::::: |||||
DB 1160 ArgAspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGluGlnGlnArgLeuAsp 1179
OY 865 CAGGACACACAGAGAGCAGCCCTGGCCAGAGAGAGTGCCTGACTTAACAGAACTGCTG 924
    ||||| |||
DB 1180 ThrAlaThrGlnGlnArgAlaGlnLeuGlnAlaGlnLeuAlaArgLeuAlaAlaAspArg 1199
OY 925 GCGGAATCTGAGCACCACCTGCACTCCACAGATCTGAAATAGCTCAA----- 972
    ||||| |||||
DB 1200 AspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGlnGlnArgLeuAspThr 1219
OY 973 CTCAGTCAGAAAGAAAGATATCATATGATTAATTTGGAAAGTTACAGAGAAATGAA 1032
    ::::: :::::
DB 1220 AlaThrGlnGlnArgAlaGlnLeuGlnAlaGlnValAlaArgLeuAlaAlaAspGlyAsp 1239
OY 1033 GAATTGAGAGAACAGTGTCTCCAGATGGAGAGTACATGAGAGCTGAAGCAAGGCTA 1092
    ||| ::::: |||
DB 1240 GluAlaArgGlnGlnLeuAlaAlaAsnAla-----GlnGlnLeuGlnGlnArgLeu 1256
OY 1093 AGGCACTGGAATAGCAACAGCAGGCGACAGCCACAGCTGCTGCTCAGCAAG 1152
    ::::: |||||
DB 1257 AspThrAlaThrGlnGlnArgAlaGlnLeuGlnAlaGlnLeuAlaArgLeuAlaAlaAsp 1276
OY 1153 CAGAACCACTTCTCTCGAGAGGCGAGAGCTGTCCGAAGAGGTGACCGCTGCGGACC 1212
    ::::: |||||
DB 1277 ArgAspGlu-----AlaArgGlnGlnLeuAlaAlaAsnAlaGlnGluGlnGln 1293
OY 1213 CAGTTACCAGCATGCCACA 1233
    ::::: |||
DB 1294 ArgLeuAspThrAlaThrGln 1300

```

```

RESULT 8
US-09-991-496-120
: Sequence 120, Application US/09991496
: Patent No. US20020169285A1
: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: APPLICANT: Campos-Nelo, Antonio
: APPLICANT: Webb, John R.
: APPLICANT: Dillon, Davin C.
: APPLICANT: Skelky, Yasir A.W.
: APPLICANT: Bhatia, Ajay
: APPLICANT: Colter, Rhea
: APPLICANT: Probst, Peter
: APPLICANT: Brannon, Mark
: TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
: TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
: FILE REFERENCE: 210121.420C9
: CURRENT APPLICATION NUMBER: US/09/991,496
: NUMBER OF SEQ ID NOS: 137
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 120
: LENGTH: 2310
: TYPE: PRT
: ORGANISM: Leishmania major and chagasi
US-09-991-496-120

Alignment Scores:
Pred. No.: 3,95e-09 Length: 2310
Score: 97 Matches: 97
Percent Similarity: 43.418 Conservative: 104
Best Local Similarity: 20.958 Mismatches: 202
Query Match: 8.35% Indels: 60
DB: Gaps: 14

US-09-502-945-1 (1-1552) x US-09-991-496-120 (1-2310)
OY 7 GATGATCCGAGAGAGCTAAACTTACTTATGAGAAAGTGTGAATTTGAG----- 57
    ::::: ::::: |||||
DB 493 GluLeuGlnGlnArgLeuAspThrAlaThrGlnGlnArgAlaGlnLeuGlnAlaArgVal 512
OY 58 -----GAATCCCAATGAGTTTGGAGAACGAC 87
    |||||
DB 513 AlaArgLeuAlaAlaAspArgSpGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGlu 532
OY 88 TTAGCTGAATATCAGAGACTTGTGAAGATCTTAAGAGCACTAAAGCATTAAGAATTT 147
    ||| ::::: |||
DB 533 LeuGlnGlnArgLeuAspThrAlaThrGlnGlnArgAlaGlnLeuGlnAlaArgValAla 552
OY 148 CTTCTGCTGCTAATACT-----TGTAACCGTGTGGTGCTTTGTTGAATGT 198
    ||||| |||||
DB 553 ArgLeuAlaAlaAsnAlaGlnGlnGlnArgLeuAspThrAlaThrGlnGlnArg 572
OY 199 GCTCAGCATGAGACTTCTTCCCAACCCACTACTAATCTTCAT-----ATGCAAGAC 252
    ||||| |||||
DB 573 AlaGlnLeuGlnGlnArgValAlaArgLeuAlaAlaAsnAlaGlnGluGlnGlnArg 592
OY 253 ATCGAAGACTGTTAAGAAAGATGACTGATGTCGTCACATTAATTCGTAAGGAGC 312
    ::::: :::::
DB 593 LeuAspThrAlaThrGlnGlnArgAlaGlnLeuGlnAlaArgValAlaArgLeuAlaAla 612
OY 313 AGCTTGGCAGATACGAGCAAGAGCAAGAGTCTTATGACAGAGTGAACAAAGTTTG 372
    ::::: |||||
DB 613 AspArgAspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGlnGlnArgLeu 632
OY 373 CAATATCTGAGAGACCAATTTTGAATAAACCAAGCTTTAATCCAGTGT----- 423
    ::::: |||
DB 633 AspThrAlaThrGlnGlnArgAlaGlnLeuGlnAlaGlnLeuAlaArgLeuAlaAlaAsp 652
OY 424 ---GACCACTTGAAGAGAGTGGAGAGGCAAGCGGAGCACTTGAATAAGAACTT--- 477
    ||||| |||||
DB 653 GlyAspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGlnGlnArgLeuAsp 672

```

```

478 ---GCATCTCAGCAAGAG-----AAAGGCCATTGAGAA 510
    |||:|||||
Db 673 ThrAlaThrGlnGlnAlaGluLeuGluAlaArgValAlaArgLeuAlaAspArg 692
    |||:|||||
QY 511 GACATGATGAAAGAAATTAACGAAAGAAAGAGATGATGAGATCAAGATGTTGATC 570
    |||:|||||
Db 693 AspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnArgLeuAspThr 712
    |||:|||||
QY 571 TTGCTCAGATATATGCCCCAATCGAGCCCGAGTGCAGAAAGTTTACAAGCAAAAGATT 630
    |||:|||||
Db 713 AlaThrGlnGlnAlaArgAlaGluLeuGluAlaGlnLeuAlaArgLeuAlaAspArg 732
    |||:|||||
QY 631 TCAGCTATTAATCACTG-----GAGGAATTCAGACCGAGCTG----- 669
    |||:|||||
Db 733 GluAlaArgGlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnArgLeuAspThrAla 752
    |||:|||||
QY 670 -----GCTTCTCGGGAATGAGTGCACAAAGGTGTGGA-----GAA 708
    |||:|||||
Db 753 ThrGlnGlnAlaArgAlaGluLeuGluAlaArgValAlaArgLeuAlaAlaAspGlu 772
    |||:|||||
QY 709 ATGGCGTATGAGCTG-----AATAAACCAACATGAGAGAG-----GATGAGCAGAA 756
    |||:|||||
Db 773 AlaArgGlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnArgLeuAspThrAlaThr 792
    |||:|||||
QY 757 AAGGAGCAGAGAGATTGAGAGCAAAACTAACAGGATCTTGAATTAAGATCAGAGAA 816
    |||:|||||
Db 793 GlnGlnArgAlaGluLeuGluAlaArgValAlaArgLeuAlaAlaAspArgAspGluAla 812
    |||:|||||
QY 817 ATAGCAAAATTTGACAAATAGAACTGGATGAAACCAACACTTGGAAACAGAGCAGAC 876
    |||:|||||
Db 813 ArgGlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnGlnArgLeuAspThrAlaThrGln 832
    |||:|||||
QY 877 AAGGACCCCGCGCAAGAGAGAGTGCCTGAGACTAACAGACGCTGGGCGAGATCGAG 936
    |||:|||||
Db 833 GlnArgAlaGluLeuGluAlaGlnLeuAlaArgLeuAlaAlaAspGlyAspGluAlaArg 852
    |||:|||||
QY 937 CACCAACTGCACTCACCAGATCTGAATAGCTCAA-----CTCAGTCAAGAA 984
    |||:|||||
Db 853 GlnGlnLeuAlaAlaAsnAlaGluGlnGlnArgLeuAlaAspThrAlaThrGlnGln 872
    |||:|||||
QY 985 AAAAGGTATCATATGATTAATTTGGAAAGTTTACAGAGAAAGAAATTAAGATTTGAGGAA 1044
    |||:|||||
Db 873 ArgAlaGluLeuGluAlaArgValAlaArgLeuAlaAlaAspArgAspGluAlaArgGln 892
    |||:|||||
QY 1045 CAGGTGTCCAGTGGGAGAGTACGATGAGCGATGAAGCAAGCTTAAGGACACTGGAT 1104
    |||:|||||
Db 893 GlnLeuAlaAlaAsnAla-----GluGluLeuGlnGlnArgLeuAlaAspThrAlaThr 909
    |||:|||||
QY 1105 AAGCAGACGAGCGCAGAGCCAGCAGCTGTGAGCTCTCAGACAAAGCAAGACAGCTT 1164
    |||:|||||
Db 910 GlnGlnArgAlaGluLeuGluAlaGlnLeuAlaArgLeuAlaAlaAspArgAspGlu--- 928
    |||:|||||
QY 1165 CTCTGAGAGAGCAGAGCTGTGCGAAGAGGTGAGCCGGCTCGCGACCTTACCCAGC 1224
    |||:|||||
Db 929 -----AlaArgGlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnArgLeuAspThr 946
    |||:|||||
QY 1225 ATGCCACAA 1233
    |||
Db 947 AlaThrGln 949
    |||

RESULT 9
; Sequence 114, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; FILE REFERENCE: 063915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
```

```

; SEQ ID NO 114
; LENGTH: 2310
; TYPE: PRF
; ORGANISM: L. major
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: AC005893.12 L6202.3
; NAME/KEY: misc_feature
; OTHER INFORMATION: 916899664
; US-09-820-843A-114

Alignment Scores:
Pred. No.: 3,95e-09 Length: 2310
Score: 228.00 Matches: 97
Percent Similarity: 43.41% Conservative: 104
Best Local Similarity: 20.95% Mismatches: 202
Query Match: 9 Indels: 60
DB: Gaps: 14

US-09-502-945-1 (1-1552) x US-09-820-843A-114 (1-2310)
QY 7 GATCATCCGAGAAAGCTPAAACTTACTTATGAGGAAAGTGTGAATTGAG----- 57
    |||:|||||
Db 493 GluLeuGlnGlnArgLeuAspThrAlaThrGlnGlnArgAlaGluLeuGluAlaArgVal 512
    |||:|||||
QY 58 -----GAATCCCAATTGAAGTTTGGAGGAGAC 87
    |||:|||||
Db 513 AlaArgLeuAlaAlaAspArgAspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGluGlu 532
    |||:|||||
QY 88 TTACCTGAAATATCAGAGAACTTGTGAAGATCTTAAAGAGCACTTAAGCATTAAGAAATTT 147
    |||:|||||
Db 533 LeuGlnGlnArgLeuAspThrAlaThrGlnGlnArgAlaGluLeuGluAlaArgValAla 552
    |||:|||||
QY 148 CTTGCTGCTGTAATCT-----TGTAACCGTGTGTGCTGCTTTGTTGAATGT 198
    |||:|||||
Db 553 ArgLeuAlaAlaAsnAlaGluGluLeuGlnGlnArgLeuAspThrAlaThrGlnGlnArg 572
    |||:|||||
QY 199 GCTGAGATGAGCTGTTCTTCCCAACCATCTAATGTTCTAT-----ATGCAGACC 252
    |||:|||||
Db 573 AlaGluLeuGluAlaGlnValAlaArgLeuAlaAlaAsnAlaGluGluLeuGlnGlnArg 592
    |||:|||||
QY 253 ATCGAAAGACTGTTTAAAGAAAGATGACTGTGTCGTGCACTTAAGTTCGTAAAGAGC 312
    |||:|||||
Db 593 LeuAspThrAlaThrGlnGlnArgAlaGluLeuGluAlaArgValAlaArgLeuAlaAla 612
    |||:|||||
QY 313 AGCTTGGCAGATACGACGAGCAAGAGCAAGCTTATGACAGGAGTGAAGAGTTTG 372
    |||:|||||
Db 613 AspArgAspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnGlnArgLeu 632
    |||:|||||
QY 373 CAATATCTGAGGAAGCAATTTTGAAGAAACCAAGCTTTAATCCAGTGT----- 423
    |||:|||||
Db 633 AspThrAlaThrGlnGlnArgAlaGluLeuGluAlaGlnLeuAlaAlaArgLeuAlaAlaAsp 652
    |||:|||||
QY 424 ---GACCAATTGAGGAAGAGCTGAGAGCGGAGCGCACTTGAAGAAAGACTT--- 477
    |||:|||||
Db 653 GlyAspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnGlnArgLeuAsp 672
    |||:|||||
QY 478 ---GCATCTCAGCAAGAG-----AAAGGCCATTGAGAA 510
    |||:|||||
Db 673 ThrAlaThrGlnGlnArgAlaGluLeuGluAlaArgValAlaArgLeuAlaAlaAspArg 692
    |||:|||||
QY 511 GACATGATGAAAGAAATTAACGAAAGAAAGAGATGATGAGATCAAGATGTTGATC 570
    |||:|||||
Db 693 AspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnArgLeuAspThr 712
    |||:|||||
QY 571 TTGCTCAGATATATGCCAATCGAGCCCGAGTGCAGAAAGTTTACAAGAAAGAAAGATT 630
    |||:|||||
Db 713 AlaThrGlnGlnArgAlaGluLeuGluAlaGlnLeuAlaArgLeuAlaAlaAspArgAsp 732
    |||:|||||
QY 631 TCAGCTATTAATCACTG-----GAGGAATTCAGACCGAGCTG----- 669
    |||:|||||
Db 733 GluAlaArgGlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnGlnArgLeuAspThrAla 752
    |||:|||||
```



```

QY 670 -----GCTTCTCGGGAATGATGTCACAAAGGTGTGCA-----GAA 708
Db 753 ThrGlnGlnArgAlaGluLeuGlnAlaArgValAlaArgLeuAlaAlaAspGlyAspGlu 772
QY 709 ATGGCGTATCAGCTG-----ATAAACCACATGAGAG-----GATGAGCGAGAA 756
Db 773 AlaArgGlnGlnLeuAlaAlaAsnAlaGluGluGlnGlnArgLeuAspThrAlaThr 792
QY 757 AAGGACACAGAGATTTCAGACAAAACCTAACAGGATCTTGAATTAAATTCAGAA 816
Db 793 GlnGlnArgAlaGluLeuGlnAlaArgValAlaArgLeuAlaAlaAspArgAspGluAla 812
QY 817 ATAGAGAAATTGAGATTGACTGATGTAAGCAACAACACTGGACAGAGACGAG 876
Db 813 ArgGlnGlnLeuAlaAlaAsnAlaGluGlnGlnArgLeuAspThrAlaThrGln 832
QY 877 AAGGACCGCTCGGCGACAGAGAGTCCGAGCTAACAGAACTGCTGGGCGCAATCTGAG 936
Db 833 GlnArgAlaGluLeuGlnAlaGlnLeuAlaArgLeuAlaAlaAspGlyAspGluAlaArg 852
QY 937 CACCAACTGCACCTCACAGATCTGAAATAGCTCAA-----CTCAGTCAAGAA 984
Db 853 GlnGlnLeuAlaAlaAsnAlaGluGlnGlnArgLeuAspThrAlaThrGlnGln 872
QY 985 AAAAGGTATACATGATGATAAATTGGGAAAGTTACAGAGAAAGAAATGAGAGAA 1044
Db 873 ArgAlaGlnLeuGlnAlaArgValAlaArgLeuAlaAlaAspArgAspGluAlaArg 892
QY 1045 CAGTGTGTCCACATGAGAGAGTACTGAGACGATGACAAAGGCTAAGGACGCTGAGT 1104
Db 893 GlnLeuAlaAlaAsnAla-----GlnGlnLeuGlnGlnArgLeuAspThrAlaThr 909
QY 1105 AAGCAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1164
Db 910 GlnGlnArgAlaGlnLeuGlnAlaGlnLeuAlaArgLeuAlaAlaAspArgAspGlu 928
QY 1165 CTCCTGGAGAGCAGACGCTGTGCGAAGAGGTGACCGGCTCGGCGACCAAGTTACCCAC 1224
Db 929 -----AlaArgGlnGlnLeuAlaAlaAsnAlaGlnGlnGlnArgLeuAspThr 946
QY 1225 ATGCCACAA 1233
Db 947 AlaThrGln 949

RESULT 10
US-09-874-923-120
: Sequence 120, Application US/09874923
: Patent No. US20020081320A1
: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: APPLICANT: Campos-Neto, Antonio
: APPLICANT: Webb, John R.
: APPLICANT: Dillon, Davin C.
: APPLICANT: Skelky, Yasir A.W.
: APPLICANT: Bhatia, Ajay
: APPLICANT: Coler, Rhea
: APPLICANT: Probst, Peter
: APPLICANT: Brannon, Mark
: TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
: TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
: FILE REFERENCE: 210121.420C8
: CURRENT APPLICATION NUMBER: US/09/874.923
: NUMBER OF SEQ ID NOS: 122
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 120
: LENGTH: 2310
: TYPE: PRT
: ORGANISM: Leishmania major and chagasi
US-09-874-923-120
Alignment Scores: 3.95e-09 Length: 2310

```

```

Score: 228.00 Matches: 97
Percent Similarity: 43.41% Conservative: 104
Best Local Similarity: 20.95% Mismatches: 202
Query Match: 8.35% Indels: 60
Db: 10 Gaps: 14

US-09-502-945-1 (1-1552) x US-09-874-923-120 (1-2310)
QY 7 GATCATCCGACGAGCTAAACCTTACTTATGAGAAAGGTGAATAG-----57
Db 493 GlnLeuGlnGlnArgLeuAspThrAlaThrGlnGlnArgAlaGluLeuGlnAlaArgVal 512
QY 58 -----GAATCCCAATTGAAAGTTTGGAGCAAGAC 87
Db 513 AlaArgLeuAlaAlaAspArgAspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGlu 532
QY 88 TTAAGCTGATATTCAGAGAACTTGTGAAGCTTAAAGACCACTAAAGCTAAAGATTT 147
Db 533 LeuGlnGlnArgLeuAspThrAlaThrGlnGlnArgAlaGluLeuGlnAlaArgValAla 552
QY 148 CTTCTGCTGCTAATACT-----TGTAACCGGTGCTGCTTGTGTAATGT 198
Db 553 ArgLeuAlaAlaAsnAlaGluGlnGlnArgLeuAspThrAlaThrGlnGlnArg 572
QY 199 GCTCAGCATGAAGCTGCTTCTCCCAACCCATCTAATGTTCTAT-----ATGCAGACC 252
Db 573 AlaGlnLeuGlnAlaGlnAlaAlaArgLeuAlaAlaAsnAlaGluGlnGlnArg 592
QY 253 ATCGAAGACTGTTAAAGAAAGAGTACTGATGCTGCACTATTTCCGTAAGAGAC 312
Db 593 LeuAspThrAlaThrGlnGlnGlnArgAlaGluLeuGlnAlaArgValAlaArgLeuAlaAla 612
QY 313 ACCTTGACGATACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 372
Db 613 AspArgAspGluAlaArgGlnGlnGlnLeuAlaAlaAsnAlaGluGlnGlnArgLeu 632
QY 373 CAATATCTGAGAGACCAATTTGAAAAAACCAAGCCTTTAATCCAGTGT-----423
Db 633 AspThrAlaThrGlnGlnGlnArgAlaGluLeuGlnAlaGlnLeuAlaArgLeuAlaAlaAsp 652
QY 424 -----GACCACTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT---477
Db 653 GlyAspGluAlaArgGlnGlnGlnLeuAlaAlaAsnAlaGluGlnGlnArgLeuAsp 672
QY 478 ---GCATCTCAGCAAGAG-----AAAAGGCCATTGAGAA 510
Db 673 ThrAlaThrGlnGlnArgAlaGluLeuGlnAlaArgValAlaArgLeuAlaAlaAspArg 692
QY 511 GACATGATGAAGAGAAATTAACGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 570
Db 693 AspGluAlaArgGlnGlnGlnLeuAlaAlaAsnAlaGluGlnGlnArgLeuAspThr 712
QY 571 TTGCTCAGAAATATGCCCACACTGGGCGGCGGAGAAAGGTTTCAAGGAAAGAT 630
Db 713 AlaThrGlnGlnArgAlaGluLeuGlnAlaGlnGlnLeuAlaArgLeuAlaAlaAspArgAsp 732
QY 631 TCAGCTATTATCAACTG-----GAGAAATTCAAAGCCAGCTG-----669
Db 733 GlnAlaArgGlnGlnLeuAlaAlaAsnAlaGluGlnGlnGlnArgLeuAspThrAla 752
QY 670 -----GCTTCTCGGGAATGATGTCACAAAGGTGTGCA-----GAA 708
Db 753 ThrGlnGlnArgAlaGluLeuGlnAlaArgValAlaArgLeuAlaAlaAspGlyAspGlu 772
QY 709 ATGGCGTATCAGCTG-----ATAAACCACATGAGAG-----GATGAGCGAGAA 756
Db 773 AlaArgGlnGlnLeuAlaAlaAsnAlaGluGlnGlnArgLeuAspThrAlaThr 792
QY 757 AAGGACACAGAGATTTCAGACAAAACCTAACAGGATCTTGAATTAAATTCAGAA 816
Db 793 GlnGlnArgAlaGluLeuGlnAlaArgValAlaArgLeuAlaAlaAspArgAspGluAla 812
QY 817 ATAGAGAAATTGAGATTGACTGATGTAAGCAACAACACTGGACAGAGACGAG 876

```

Db	813	ArgInGInGInLeuAlaAlaAsnAlaGluGluLeuGInGInIhrArgLeuAspIhrAlaThrGIn	832
QY	877	AAGCAGCCCTGGCCAGAGAGAGAGTGGCTGGAGACTAACAGAACTGCTGGCGCAATCTGAG	936
Db	833	GInArgAlaGluLeuGInGluAlaGInLeuAlaArgLeuAlaIAspGlyAspGluAlaArg	852
QY	937	CACCAACTGCACCTCCACCACATCTGAAATAGCTCAA-----CTCACTCAGAA	984
Db	853	GInGInLeuAlaAlaAsnAlaGluGInGInGInIhrArgLeuAspIhrAlaThrGInGIn	872
QY	985	AAAGGATACATATGATTAATTTGGGAAGTTACAGAGAAAGAAATGCAAGATTGGAGSAA	1044
Db	873	ArgAlaGluLeuGInGluAlaIhrGlyAlaIaArgLeuAlaIAspArgSpGluAlaArgGIn	892
QY	1045	CAGGTGTCACGATGGGAGAGATGACATGACAGACATGACGAAGAAGCCTTAAGCGAGCTGAT	1104
Db	893	GInLeuAlaAlaAsnAla-----GluGluLeuGInGInIhrArgLeuAspIhrAlaThr	909
QY	1105	AAGCAGCCAGCGCCACAGCCAGCCAGCTGGTGCAGCTCTTCAGCAAGACGAAACCACTT	1166
Db	910	GInGInIhrArgAlaGluLeuGInGluAlaGInLeuAlaArgLeuAlaIAspArgSpGlu---	928
QY	1165	CTCTCGAGAGGACGAGACCTGTGGGAAGAGGTGGACCGGCTCGGAGCCAGTTACCCAGC	1222
Db	929	-----AlaArgGInGInLeuAlaAlaAsnAlaGluGluLeuGInIhrArgLeuAspIhr	946
QY	1225	ATGCGACAA 1233	
Db	947	AlaThrGIn 949	

```

Db 731 ArgArgArgGluAspAsnLeuSnaSpSerSerGlnGlnLeuValGlnGluLeuLeu 750
QY 964 ATAGCT-----CAACTGAGTCAAGAAAAAGGTATATGATATGATTAATGGAAAGTTA 1017
Db 751 MetAlaMetGlnLysValLysGlnGlnLeuGlnSerMetLysAlaLysLeuSerSerThr 770
QY 1018 CAGAGAGAAGATGAGAAATGGAGAAACAGTGTCTGCAGCATGGAGACAGTACATGAGACG 1077
Db 771 GlnGlnSerLeuAlaGlnLysGln-----ThrHisLeuThr 782
QY 1078 -----ATGACGAAAGCTTAAGGACGCTGATAGCACACGACGACGACGACGACG 1131
Db 783 AsnLeuArgAlaGlnLysArgGlnHisLeuGlnGlnValLeuGlnMetLysGlnGlnAla 802
QY 1132 CAGGCTGACCTCTCTGACGACGACG-----AACACACTTCTCTCTGAGAGGACGACG--- 1182
Db 803 LeuLeuAlaAlaIleSerGlnLysAspAlaAlaIleAlaLeuLeuLeuSerSerSer 822
QY 1183 -----CTGTGCGAAGAGGCTGACGCGCTGCGGACG 1212
Db 823 LysLysLysThrGlnGlnGlnValAlaAlaAlaLeuLysArgGlnLysAspArgLeuValGln 842
QY 1213 CAGTTACCCAGCATGCGCACAAATCTGATTGCTGACCTGATGGAACAGAGTGAATTAATG 1272
Db 843 GlnLeuLysGlnGlnThrGln-----AsnArgMetLysLeuMet 855
QY 1273 -----ATTACAAACAG 1284
Db 856 AlaAspAsnTyrGlnAsp 861

RESULT 12
US-09-820-843A-73
; Sequence 73, Application US/09820843A
; Publication No. US2003003963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 73
; LENGTH: 1805
; TYPE: PRT
; ORGANISM: M. genitalium
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: g1|1045905
US-09-820-843A-73

Alignment Scores:
Pred. No.: 6,86e-09 Length: 1805
Score: 224.50 Matches: 98
Percent Similarity: 43.10% Conservative: 83
Best Local Similarity: 23.33% Mismatches: 166
Query Match: 8.22% Indels: 73
DB: 9 Gaps: 15

US-09-502-945-1 (1-1552) x US-09-820-843A-73 (1-1805)
QY 16 GAGAGCTAAACTTACTTATGAGAAAAGTGAATTGAGAAATCCCAATTGAAGTTT 75
Db 499 GlnGlnLeuTyrLeuValLysLysGlnLysGlnAspGlnLysGlnLysLeuLeuPhe 514
QY 76 TTGAGAGCACTTACGCTGAATATACAGAACTTGAAGATCTTAAAGACAACAACTAAG 135
Db 515 PheGlnLysGlnLeuLysGlnLysGlnAlaAspPheGlnLysn---GlnLeuGlnAlaLys 533
QY 136 CATAAAGATTTCTTCTGCGCTGCTAATACTTGTAAACCGCTTGCGTGTCTTTGTTGAAA 195

```

```

Db 534 GlnGlnGlnLeuPheGlnLysAlaLysHisAlaLeuLysArg-----SerPheIleLys 550
QY 196 TGTGTCACATGACATGCTGTCTTCCCAAAACCATTATGATATGATATGACAGACATC 255
Db 551 LeuGlnAspLysGlnLysAspLeu-----AsnThrLysAlaGlnGlnIle 565
QY 256 GAAAGACTGGTTAAAGAAAGATGACTTGATGCTGCACTATTCCTCGTAAGGACAGC 315
Db 566 Ala-----AsnGlnPheSerGlnLeuLysThrLysPheLysSerLys 578
QY 316 TTGGCAGATACGACGCAAGAGACAGACAGTCTTATGACAGAGTGAACAAGTTTGCA 375
Db 579 SerAlaAspPheGlnLeuMetLeuGlnAsnGlnTyrGlnAsnLeuGlnGlnGlnLysGln 598
QY 376 -----ATATCTGAGGAAGCAATTTGAAAAAACCAAGCCTTAATC-----CAGTGT 423
Db 599 LysLeuPheGlnGlnLysArgThrGlnPheGlnLysArgAsnAlaAlaValLeuSerAsnAla 618
QY 424 GACCACTTGAAGAGAGAGCTGAGAGGACGACGACGACGCTTGAAGAAAGCACTTGATCT 483
Db 619 GlnGlnLysArgGlnGlnLeuLeuGlnGlnLysGlnLysGlnLysPheLysPheLysSer 638
QY 484 CAGCAAGAAAAAGGCCATTTGAGAAAGACATGATGAAAAAG-----GAAATTAAG 534
Db 639 PheGlnGlnGlnLysArgLeuLysGlnGlnLysGlnLysGlnLysGlnLysValAlaSerValGln 658
QY 535 AAAGAAAGGAGTACATGAGGATCAAGATG-----TTGATCTTG 573
Db 659 LysGlnLysGlnLysLeuLysLysLysGlnAspPheSerGlnThrSerLeuAsnAla 678
QY 574 TCTCAGAAATATGCCCCA-----CTGAGGCGCCAGGTGGAAG 612
Db 679 SerLysAsnLeuValGlnLysArgGlnMetAlaIleLysPheLysGlnLysGlnIleGlnAla 698
QY 613 GTTACAAAGAAAAAGTTTACAGCTATTAATCAACTGAGGAAATGAAACGACGAGTCT 672
Db 699 ThrGlnLysGlnLeuLeuAsnAspValAsnAsnAlaGlnValIleGlnAlaAspLeuAla 718
QY 673 TCTCGGAAGATGATGTCACAAAGGTGTGTGGAAGATCGCTATGAGTGAATAAACC 732
Db 719 -----GlnLeuAsnGlnSerLeuAsnGlnGln 727
QY 733 AACATGAGAGAGATGAGCAGCAAAAAGACACAGACGACTTACAGCAAAAACCTAAGCAG 792
Db 728 ArgSerGlnLeuGlnAsnAlaLysGlnArgIleAlaAspPheHisAsnAspSerLeuLys 747
QY 793 GATCTTGAATTAAGATACAGAAATGAGAAATGAGATGAGATGAGT-----843
Db 748 LysLeuAsnGlnLysGlnLeuSerLeuGlnLysArgLeuGlnGlnLeuGlnThrLeuGln 767
QY 844 ---GAAAGCAACACACACTTGGAAACAGAGCAGACAGAGGACCCCTGGCCAGAGAGAG 900
Db 768 AlaAsnGlnLysHisSerTyrGlnAsnGln-----AlaTyrPheGlnGlnGln 784
QY 901 TGCCTGAGACTAAGACACTGCTGGCGCAATCTGACACACCACTGACACTGACCAATCT 960
Db 785 -----LeuAspLysLeuAsnArgGlnLysGlnAlaPheLeuAsnLeuArgLysLys 801
QY 961 GAAATGCTCAACTGCTCAGCAAGAAAAGGTATACATGATGATAA-----1005
Db 802 GlnThrMetGlnValAspAlaIleLysGlnArgLeuSerAspLysHisGlnAlaLeuAsn 821
QY 1006 -----TTGGGAAAGTTTACAGAGAAAGAAATGAAAGATTGAGAGAACAGTGTCCAGCAT 1059
Db 822 MetGlnGlnAlaGlnLeuAsnArgLysThrHisGlnLysLeuAsnAlaPheLeuAsnHis 841
QY 1060 GGGAGAGTACATGACAGCATGGAAGCAAGAGCTTAAGGACGACCTGATTAAGACAGCCAGGCC 1119
Db 842 AspAlaAspGlnLysSerLeuGlnAspGlnLeu-----Ala 853
QY 1120 ACAGCCGACAGCTGCTGAGCTCTCAGCAGCAAGAGAACAGACAGTCTCTCGAGAGGAGC 1179
Db 854 ThrValLysGlnThrGlnLysLeuIleAspLeuGlnArgSerAlaLeuLeuGlnLysGln 873

```

RESULT 13  
 US-09-815-242-10617  
 : Sequence 10617, Application US/09815242  
 : Patent No. US20020061569A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Haselbeck, Robert  
 : APPLICANT: Ohlsen, Karl L.  
 : APPLICANT: Zyskind, Judith W.  
 : APPLICANT: Wall, Daniel  
 : APPLICANT: Trawick, John D.  
 : APPLICANT: Carr, Grant J.  
 : APPLICANT: Yamamoto, Robert T.  
 : APPLICANT: Xu, H. Howard  
 : TITLE OF INVENTION: Identification of Essential Genes in  
 : TITLE OF INVENTION: Prokaryotes  
 : FILE REFERENCE: EITRA.011A  
 : CURRENT APPLICATION NUMBER: US/09/815,242  
 : CURRENT FILING DATE: 2001-03-21  
 : PRIOR APPLICATION NUMBER: 60/191,078  
 : PRIOR FILING DATE: 2000-03-21  
 : PRIOR APPLICATION NUMBER: 60/206,848  
 : PRIOR FILING DATE: 2000-05-23  
 : PRIOR APPLICATION NUMBER: 60/207,727  
 : PRIOR FILING DATE: 2000-05-26  
 : PRIOR APPLICATION NUMBER: 60/242,578  
 : PRIOR FILING DATE: 2000-10-23  
 : PRIOR APPLICATION NUMBER: 60/253,625  
 : PRIOR FILING DATE: 2000-11-27  
 : PRIOR APPLICATION NUMBER: 60/257,931  
 : PRIOR FILING DATE: 2000-12-22  
 : PRIOR APPLICATION NUMBER: 60/269,308  
 : PRIOR FILING DATE: 2001-02-16  
 : NUMBER OF SEQ. ID NOS: 14110  
 : SOFTWARE: FastSeq for Windows Version 4.0  
 : SEQ ID NO 10617  
 : LENGTH: 1045  
 : TYPE: PRN  
 : ORGANISM: Enterococcus faecalis  
 US-09-815-242-10617

Alignment Scores:  
 Pred. No.: 8.68e-09 Length: 1045  
 Score: 222.50 Matches: 109  
 Percent Similarity: 40.73% Conservative: 93  
 Best Local Similarity: 21.98% Mismatches: 159  
 Query Match: 8.14% Indels: 135  
 DB: 10 Gaps: 19

US-09-502-945-1 (1-1552) x US-09-815-242-10617 (1-1045)

QY 10 GCATCCGAGAGCTAAACTACTTATGAGAGAAAGTGTGAATTTGAGGATCCCAATTG 69  
 Db 263 AATATGAGAGAGCTAAACTACTTATGAGAGAAAGTGTGAATTTGAGGATCCCAATTG 69  
 QY 70 AAGTTTGGAGAGAGCTTA-----GCTGATATGAGAGACTTGTGAAGATCTTAA 123  
 Db 283 GATGATATGAGAGAGCTTA-----GCTGATATGAGAGACTTGTGAAGATCTTAA 123  
 QY 124 GAGCACTAAAGCATTAAGAAATTTCTTGCTGCTCTAATCTTAAACCGTTGGTGGT 183  
 Db 303 GATGATATGAGAGAGCTTA-----GCTGATATGAGAGACTTGTGAAGATCTTAA 123  
 QY 184 CTTGTTGAATGTGCTGAGCATGAAGCTTTCTTCCCAACCCATTAATGTTGAT 243  
 Db 308 CTTGTTGAATGTGCTGAGCATGAAGCTTTCTTCCCAACCCATTAATGTTGAT 243  
 QY 244 ATGACAGCATGAAGAGCTGTTAAAGAAAGAGAT-----GACTTGATGTGCACTA 297  
 Db 323 ATGACAGCATGAAGAGCTGTTAAAGAAAGAGAT-----GACTTGATGTGCACTA 297  
 QY 298 GTTTCGTAAGAGAGCTGTTAAAGAAAGAGAT-----GACTTGATGTGCACTA 297  
 Db 342 GTTTCGTAAGAGAGCTGTTAAAGAAAGAGAT-----GACTTGATGTGCACTA 297

Db 343 MetGluThrGlnGlnAlaLeuThrAspTrpGlnAlaIleMetSerGluLeuGluGln 362  
 QY 331 -----CAAGAGAGAGAGAGTCT 348  
 Db 363 GlnProLeuIleAlaGlnLysGlnLysArgLeuGlnThrIleGlnArgGlnLeuProGln 382  
 QY 349 TATGACAGCTGAAG 402  
 Db 383 TATGACAGCTGAAG 402  
 QY 403 ACCAAGCTTAAATTCAGTGTGACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 462  
 Db 402 -----AlaIleGlnLysGluThrGluSer 409  
 QY 463 CTGAAAG 522  
 Db 410 CysGlnGlnGlnLysIleThrLeuAlaSprLysValAlaThrAlaLysGlnPheIleGln 429  
 QY 523 AAGCAATATGAG 570  
 Db 430 GlnGluGluThrLeuGlnLysAlaAspPheLysSerSerValAlaAspHisTrpGln 449  
 QY 571 -----TGTCTCAGATATG 585  
 Db 450 AspPheValGluArgTrpGlnLysAsnGlnLysAlaTrpGlnLysIleSerGlnAsnGln 469  
 QY 586 GCCCACTGAG 636  
 Db 470 ValGluLeuHisGluLeuThrGlnThrPheAlaValGluGlnGlnLysSerAlaGln 489  
 QY 637 ATTAATCACTGAG 675  
 Db 490 GlnAlaLysLeuGlnThrLysSerGlnThrAlaSerLeuGlnIleGlnArgLeuSer 509  
 QY 676 -----CGCAATGATGTCACAAAGGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAG 726  
 Db 510 LeuLeuLeuGlnGlnLysGlnLysProCysArgValCysGlnSerLeuGlnHisProLysGln 529  
 QY 727 AAACCAATGAG 771  
 Db 530 GlnThrHisGlnGlnLysValSerLeuGlnLysIleAspGlnAlaGlnArgGlnLeuThrGln 549  
 QY 772 TTCAGAGCAAAATTAACAG 810  
 Db 550 ValGluThrValGlnArgPheThrGlnThrLeuSerAlaLeuGlnAlaGlnLysGln 569  
 QY 811 CAGCAATATGAG 870  
 Db 570 GlnLysGlnSerGlnLeuGln-----GlnGlnGlnAlaAlaLysThrGlnGlnGln 586  
 QY 871 CAGCAAG 930  
 Db 587 GlnGlnLeuAlaAlaGlnPheAlaAspLeuGlnLeuProLeuThrGlnLeuThrPheSer 906  
 QY 931 TCTGAGCAGCAATGAG 978  
 Db 607 -----GlnValThrProAlaGlnLysLeuAlaGlnLysGlnLeuAla 621  
 QY 979 CAGCAAG 1029  
 Db 622 LysGlnLysGlnGlnIleAlaGlnLysLeuThrGlnLysLeuValGlnLysAspArgLeu 641  
 QY 1030 GAAGAATTTGAG 1071  
 Db 642 AlaGlnLeuGlnGlnLysValAlaGlnLysSerGlnArgPheGlnValLeuArgGlnGln 661  
 QY 1072 ---GAGAGATGAG 1128  
 Db 662 ValGluThrMetGlnGlnSerLeuGlnArgIleThrIleGlnGlnGlnMetIleAlaSer 681  
 QY 1129 CAGCTGTG-----CAGCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1170  
 Db 682 GlnLeuLeuAspAlaThrValThrThrGlnGlnMetThrLysGlnGlnAlaLeuLeuGln 701

QY 1171 GAGAGGACAGACCTGTCGGAA---GAGGTGACCGGCTGCGGACCCAG 1215  
||| ||| ||| ::| ::| |||||  
Db 702 GUGluLeuSerAlaPheGluArgGlnLysGluAsnValThrThrGln 717

RESULT 14  
US-10-017-216-2  
; Sequence 2, Application US/10017216  
; Patent No. US20020160483A1

```

: APPLICANT: KAPPELLER-LIBERMAN, Rosana
: TITLE OF INVENTION: 13245, A No. US20020160483A1 Human Mitochondrial Dystrophy Type Protein
: TITLE OF INVENTION: Kinase and Uses Therefor
: FILE REFERENCE: 10147-5701
: CURRENT APPLICATION NUMBER: US/10/017,216
: CURRENT FILING DATE: 2001-10-23
: PRIOR APPLICATION NUMBER: US 60/242,429
: PRIOR FILING DATE: 2000-10-23
: NUMBER OF SEQ ID NOS: 7
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2
: LENGTH: 2053
: TYPE: PRT
: ORGANISM: Homo sapiens
: OS-10-017-216-2

```

Alignment Scores:	
Pred. No.:	1.53e-08
Score:	220.00
Percent Similarity:	42.27%
Best Local Similarity:	20.45%
Query Match:	8.05%
DB:	9
	Gaps:
	10
	205
	Length:
	Matches:
	Conservative:
	Mismatches:
	Indels:
	60
	10

US-09-502-945-1 (1-1552) x US-10-017-216-2 (1-2053)

QY	7	GATCATCCCGAAGACTTAAAGCACTTACTATGAGAAAGCTGCAATTGAC-----	57
Db	640	GIUleugInGIuIySleugInLyAlAlAluYsgIuIaGIaGIuIaYgIuIeugInIyS	655
QY	58	-----GAATCCCAATTAAGTTTTGAGAAAGCAAGCACTTAGCTGATATACAGA	105
Db	660	leugInAsnaArgIuAspSerSerGIuIyIleArgIySlySleuValGIuIaGIuIu	675
QY	106	ACTGTGAAAGTCTTAAAGCAACATAAGCATTAAGCAATTTCTTGGCTGCTAAATACT	165
Db	680	ArgIArgInISerIeugInuAsnLyValIySArgIyIeugInIhImeGIuIaArgIuI--	695
QY	166	TGTAAACGCTGTGGTGT-----	183
Db	699	---AsnaArgIeIySAspAspIleGIInIhIySergInGIleGIInImeAlAsp	711
QY	184	-----CTTGTCTTGAATGTGCTCAGCATGACAGCTGTCTTCCCAAGCCATATAAT	231
Db	718	LySIIeIeugInIeugInGIuIySInIArgIuIaGIaInISerIaGIaInISleugInI	733
QY	238	GTTATATGCGACCAATCGAAAGACTGTTAAAGAAAGATGACTTGATGTGCACTA	299
Db	738	ValnISleuYsgInIySgInGIuInISyTgIuGIuIySIleIySValIeAspAsn--	756
QY	298	GTTTCCGGAAGGACCACTGCGAGATTAAGCAGCAAGAGAAAGCAAGCTTATGACAG	355
Db	757	---GIInIeIySIySAspIeIuAlAspIyS-----GIuIhI	767
QY	358	GTCGAACAAGTTTGTCAAAATATCTGAGAGAACCAATTTGAAAAACCAAGCTTATATC	417
Db	768	leugInuAsnImeTGTInIArgInISgInGIuIaInIaInISgIySgIySIleIeugSer	783
QY	418	CAGTGTACCAAGTTGAGAAAGCACTGGAGAGCAGGCGAGCACTTGAAGAAAGCAATT	477
Db	788	GIuGIuInLySAlaMeTIIeAsnAlaMeTAspSerLySIleArgSerIeugInGIaInArgIle	807
QY	478	GCAATCTAGCAAGAGAAAGGCCAATT-----	504

Db 808 ValGluLeuSerGlnAlaAsnIleuAlaIAsnSerSerLeuPheThrGlnArgAsn 827  
Oy 505 -----GAGAAAGCATATGATGAAAAAGAAATTAACGAAAGAAAGGAGTACATGGATCA 558  
Db 828 MetIysAlaGlnGluGluMetIleSerGluLeuArgGlnGlnIlysPheTyrLeuGluThr 847  
Oy 559 AAGATGTTGACCTTGCTCCAGAAATATTGCCCAACCTGAGGCGCCAGGTGGAAAGATTACA 618  
Db 848 GlnAlaGluIlysLeuGlnAlaGlnIAsnArgIlysLeuGlnGlnIleuGlnIlysIleSer 867  
Oy 619 AAGGAAAGATTTACGCTATTATATCACTGGAGGAAATTCAAAGCCAGCTGGCTTCGG 678  
Db 868 HisGlnAsnHisSerAspIlyAsnArgIleuGluLeuGlnIlyThrArgLeuAlaArgGluVal 887  
Oy 679 GAATGAGATGTCACAAAGGTGTGTGGAAATGCGCTATCAGCTGAAATTAACCAACATG 738  
Db 888 SerLeuGlnHisGlnGluGlnIlyLeuGlnIleuLeuIlyAsnArgGlnIleuThrGlnLeuGlnIleu 907  
Oy 739 GAGAAAGATGAAGCCAGAAAGAGAG-----CACAGAGATTTCAGAGCAAAACCTAAC 789  
Db 908 SerLeuGlnGlnIlyArgGlnSerGlnIleuThrAlaLeuGlnAlaAlaArgAlaAlaLeuGln 927  
Oy 790 AGGCACTTGAATTAAGATCAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCA 849  
Db 928 SerGlnIleuAlaArgGlnAlaIlyThrGlnIleuGlnIlyThrIleAlaGlnAlaGlnIlyGln 947  
Oy 850 AAACAA-----CACTTGGACAGAGACAGCAGCAAGAGCCAGCCCTGGCCAGAGAG 897  
Db 948 IleGlnAlaLeuThrAlaHisIlyArgAspGlnIleGlnAlaIlyIysPheAspAlaLeuAlaArgAsn 967  
Oy 898 GAGTGCCTGAGACGTACAGAACTGCTGGCGCAATCTGTAGCACAACCTCAGCTCACCAGCA 957  
Db 968 SerCysThrValIleThrHisPleu-----GluGlnGlnIleuAsnGlnIleuThr 983  
Oy 958 TCTGAATGATCTCAACTCAGTCAGCAAAAAAGGTATACATATGATTAATTTGGAAAGTTA 1017  
Db 984 GluAspAsnAlaGlnIleuAsnAsnGlnHisnIlePheTyrLeuSerIlyGlnIleuAspGlnIle 1003  
Oy 1018 CAGACAACAATGAAGAATTGGAGACAGACAGTGTGTCCAGCATGGAGAGGTACATGAGACG 1077  
Db 1004 SerGlnIlyAlaAsnAspGlnIleValGlnIleuArgSerGlnValAspHisIleAlaArgGln 1023  
Oy 1078 ATGAAGCAAAAGCTTAAGCAGCTGGATGAAGACAGCCAGCCACAGCCACAGCTGGTG 1137  
Db 1024 IleThrGlnIlyArgGluMetGlnIleuThrSerGlnIlyGln---ThrMetGlnAlaIleuIly 1042  
Oy 1138 CAGCTCTCAGCAGACAGACACCCTTCTCCTGGAGAGCAGACGCTTCGGAAGAGGTG 1197  
Db 1043 ThrThrCysThrMetLeuGlnGlnIleuValMetAspLeuGlnAlaIleuAsnAspIleu 1062

RESULT 15  
US-09-864-761-47959  
Sequence 47959, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263,6  
PRIOR FILING DATE: 2000-10-04

1 APPLICANT: Hanzel, David K.  
 2 APPLICANT: Chen, Wensheng  
 3 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 4 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
 5 FILE REFERENCE: Aeomica-x-1  
 6 CURRENT APPLICATION NUMBER: US/09/864,761  
 7 CURRENT FILING DATE: 2001-05-23  
 8 PRIOR APPLICATION NUMBER: US 60/180,312  
 9 PRIOR FILING DATE: 2000-02-04  
 10 PRIOR APPLICATION NUMBER: US 60/207,456  
 11 PRIOR FILING DATE: 2000-05-26  
 12 PRIOR APPLICATION NUMBER: US 09/632,366  
 13 PRIOR FILING DATE: 2000-08-03  
 14 PRIOR APPLICATION NUMBER: GB 24263.6  
 15 PRIOR FILING DATE: 2000-10-04



Db 594 ArgIleasn-----ProThrThrValLysMetLys 603  
OY 1267 TAATGAAATTACAAGAGATATTACATTCATCTGGTTAGACTTAATATG 1317  
          ::: ||| |||||::: ||| |||  
Db 604 SerSerValPheaspGlnaspLysThrPheValAlaGlnThrLeuGlnMet 620

Search completed: March 21, 2003, 13:55:53  
Job time : 47.6182 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: March 21, 2003, 12:34:32 ; Search time 12.0233 Seconds  
(without alignments)  
7595.986 Million cell updates/sec

Title: US-09-502-945-1

Perfect score: 2732

Sequence: 1 cttctgtagctccgagaa.....aaatgaactttaagaaga 1552.

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 262574 segs, 29422922 residues  
Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame-n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool/US0502945/Runat\_14032003\_101059\_19145/app.query.fast\_1.10979  
-DB=Issued\_Patents\_AA -QFMT=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUPMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0502945.eccn\_1.1.80\_@runat\_14032003\_101059\_19145 -KCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	233.5	8.5	631	4	US-08-477-831C-11 Sequence 11, Appl1
2	232.5	8.5	976	4	US-09-104-324B-4 Sequence 6, Appl1
3	232.5	8.5	2482	1	US-08-328-254-6 Sequence 6, Appl1
4	232	8.4	606	4	US-08-477-831C-2 Sequence 2, Appl1
5	230.5	8.4	1939	4	US-09-310-187A-1 Sequence 1, Appl1
6	227.5	8.3	3248	5	US-08-353-700-1 Sequence 1, Appl1
7	227.5	8.3	3248	5	PCT-US95-16216-1 Sequence 1, Appl1
8	224	8.2	1388	4	US-09-572-191-2 Sequence 2, Appl1
9	224	8.2	1388	4	US-09-723-262-2 Sequence 2, Appl1
10	224	8.2	1388	4	US-09-723-219-2 Sequence 2, Appl1
11	223	8.2	1898	1	US-08-056-200-94 Sequence 94, Appl1
12	223	8.2	1898	2	US-08-800-644-94 Sequence 94, Appl1

13	221.5	8.1	1388	2	US-08-685-576-1 Sequence 1, Appl1
14	221	8.1	955	1	US-08-006-676B-1 Sequence 1, Appl1
15	221	8.1	955	1	US-08-282-845-2 Sequence 2, Appl1
16	221	8.1	955	5	US-08-428-414A-3 Sequence 3, Appl1
17	221	8.1	955	5	PCT-US94-00324-1 Sequence 1, Appl1
18	220.5	8.1	1886	4	US-08-938-105-3 Sequence 3, Appl1
19	218.5	8.0	2101	1	US-08-466-390-4 Sequence 4, Appl1
20	218.5	8.0	2101	1	US-08-470-950-4 Sequence 4, Appl1
21	218.5	8.0	2101	1	US-08-467-781-4 Sequence 4, Appl1
22	218.5	8.0	2101	1	US-08-195-487-4 Sequence 4, Appl1
23	218.5	8.0	2101	2	US-08-483-924-4 Sequence 4, Appl1
24	218.5	8.0	2101	4	US-09-452-294-1 Sequence 1, Appl1
25	218.5	8.0	2101	5	PCT-US93-06160-4 Sequence 4, Appl1
26	218	8.0	1093	5	PCT-US93-03077-1 Sequence 1, Appl1
27	217.5	8.0	1388	2	US-08-685-576-4 Sequence 4, Appl1
28	217	7.9	1312	2	US-08-687-080-51 Sequence 51, Appl1
29	216.5	7.9	885	2	US-08-533-306A-4 Sequence 4, Appl1
30	216.5	7.9	885	2	US-08-742-923A-4 Sequence 4, Appl1
31	216	7.9	1312	2	US-08-592-126-148 Sequence 148, App
32	215.5	7.9	1354	3	US-08-685-871-2 Sequence 2, Appl1
33	215	7.9	514	2	US-08-960-022-14 Sequence 14, Appl1
34	214.5	7.9	576	2	US-08-533-306A-2 Sequence 2, Appl1
35	214.5	7.9	576	2	US-08-742-923A-2 Sequence 2, Appl1
36	214.5	7.9	816	2	US-08-533-306A-6 Sequence 6, Appl1
37	214.5	7.9	816	2	US-08-742-923A-6 Sequence 6, Appl1
38	213	7.8	900	2	US-08-630-822A-62 Sequence 62, Appl1
39	213	7.8	900	2	US-09-005-069-62 Sequence 62, Appl1
40	213	7.8	900	4	US-09-171-156A-21 Sequence 21, Appl1
41	209.5	7.7	1066	4	US-09-541-782-8 Sequence 8, Appl1
42	209.5	7.7	1066	4	US-09-723-820-8 Sequence 8, Appl1
43	208	7.6	443	2	US-08-795-475-6 Sequence 6, Appl1
44	201.5	7.4	1162	2	US-08-728-323A-2 Sequence 2, Appl1
45	201.5	7.4	1162	4	US-09-298-568-2 Sequence 2, Appl1

#### ALIGNMENTS

RESULT 1  
US-08-477-831C-11  
; Sequence 11, Application US/08477831C  
; Patent No. 6429291  
; GENERAL INFORMATION:  
; APPLICANT: TURLEY, EVA A.  
; APPLICANT: SHUMEN, ZHANG  
; APPLICANT: ENTWISTLE, JOYCELYN  
; TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & NEAVE  
; STREET: 1251 AVENUE OF THE AMERICAS  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 10020-1104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Rel. #1.0, ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/477, 831C  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PIERRI, MARGARET A.  
; REGISTRATION NUMBER: 30,709  
; REFERENCE/DOCKET NUMBER: SIM-10  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-596-9000  
; TELEFAX: 212-596-9090  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 631 amino acids  
;

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: /desc = "RHAMM I-2a"
US-08-477-831C-11

```

```

Alignment Scores:
Pred. No.: 1,286-13 Length: 631
Score: 233.50 Matches: 126
Percent Similarity: 41.47% Conservative: 100
Best Local Similarity: 23.12% Mismatches: 170
Query Match: 8.55% Indels: 149
DB: Gaps: 28

```

US-09-502-945-1 (1-1552) x US-08-477-831C-11 (1-631)

```

QY 16 GGAAGCTAAACTTACTTATGAGAAAGTGAATGAGAA----- 60
DB 58 GIULYSGLULYSIL-----ASPGIULYSCYSGLUTHTGILULYSLEULEUGLUTYRILE 75
QY 61 -----TCCCAATGAGATTTTGGAGACGACTTACGTAAT 99
DB 76 GINGIULIESERCYSALASERASPCINVALIGIULYSCYSLYSVALASPILEAAGINLEU 95
QY 100 CAGAGAACTGTGAAGCTTAAAGAGCAACTAAAGCATTAAGATTTCTT----- 150
DB 96 GIU-----GIULASPLEULYSGLU-----LYSASPRIGIULILEUSERLEULYS 110
QY 151 -----CTGGCTGCTAATACT-----TGTACCGTGTGGTGCTTGTGTTGAATGT 198
DB 111 GINSEULEUGLUGLULASINLEUPHESERLYSGINLEGLULASPLEULHVALLYSCYS 130
QY 159 -----GTCACGATGAGCTGTCTTCCCAAAACCAT----- 231
DB 131 GINLEULEUGLUTHTGICLARGSPASINLEUVALISERLYSASPRIGIULARGIAGIUTHT 150
QY 232 ACTAATGTTCATATGCAGACATC-----GAAAGCTGTGTTAAAGAAAGAGTAC----- 282
DB 151 LEUSERIALIAGIUMETGINLEULEUTHRGIUARGLEUALALEUGIULARGINGIULUTYRGIU 170
QY 283 ---TGATGTCTGCACACTAGTTCCGTAAAGAGCAGCTTGCGAGATACGACAGAAAGAGA 339
DB 171 LYSLEUGLINGIULYSGIULINLEUGINSEULEULEU-----GLINGIULIULYSGLIU 188
QY 340 GCAAGTGTCTTATGACAGAGTGGAACAAGTTTGCATAATCTGAGAGAACCAATTTTGA 399
DB 189 LEUSERIALIARGLEUGIN---GLINGIULEUCYSSERPHGGINGLIUMETHERSERGIU 207
QY 400 AAACCC-----AAGCTTTAATCCAGTGTGACCAGTGAAGAGAG 438
DB 208 LYSASNVALPHELYSGLUGIULYULEULYSLEUALALEUALAGIULLEUASPRILAVAGINGIN 227
QY 439 GAGCTGAGAGAGCGAGCGGAGGACTTGAAGAAAGAACTTGCAATCTCAGCAAGAGAAAG 498
DB 228 LYS---GIUGIUGINSEULEULARGLEUVALIYLSGINLEUGIUGIULARGIULYSERHTH 246
QY 499 GCC-----ATTGAGAAAGACATGATGATAAAGAAAGTAACGAAAGAGAG 543
DB 247 ALAAGIUGINLEUTHRARGLEUASPRASINLEUARGIULYSGIULVALIGIULEUENIULYS 266
QY 544 GAGTACATGGCATCAAGATGTTGATCTTG-----TCT 576
DB 267 HISILEALIALHIALIAGIINALEILEULEULEIAGIINGIULYTYRASNSPHTHRIA 286
QY 577 CAGATATATGCCCACTGGAGGCCGAGGTGGAAGGTTTACAAAGAAAGATTTCCAGCT 636
DB 287 GINSEULEUARGASPRVALIHTHRIAGINLEUGINSEULVALIGINGIULYTYRASNSPHTH 306
QY 637 ATTATCAACTGAGAGAAATTCAAAGCCAGCTGTCTCGGGA----- 681
DB 307 ALAIGINSEULEUARGASPRVALIHTHRIAGINLEUGINSEULGIUGINGIULYTYRASNSP 326

```

```

QY 682 -----ATGATGTCAACAAGCTGTGTGCAANAATGCCCTATACCTGAAT 726
DB 327 THRIALAGINSEULEUARGASPRVALIHTHRIAGINLEUGINSEULGIULYTYRASNSP 346
QY 727 AAACCAACATGTGAG---AAGGAT-----GAGCAGAAAGAGGACAGACA 768
DB 347 ASPHTHRIALAGINSEULEUARGASPRVALIHTHRIAGINLEUGINSEULVALIGIULYTYR 366
QY 769 GAGTTCAGAGCAAAAACATAACAGGATCTT----- 798
DB 367 ASNASPTHRIALAGINSEULEUARGASPRVALIHTHRIAGINLEUGINSEULYTYRYSER 386
QY 799 -----GAAATTHAAGATCAGCAAAATAGCAAAATGAGATAGAA----- 837
DB 387 THRLEULYSGLIULILEGLIULASPLEULYSLEUGIULANLEUTHRLEUGINGIULYSVALA 406
QY 838 ---CTGATGAAAGCAACAACATCTGGAACAGGAG----- 870
DB 407 METALIGIULYSERVALIGIULASPRVALIGINGINGINILEUTHRIALAGIUSERHTHRSN 426
QY 871 -----CAGCAAGAGCAGCCCTGGCCAGAGAGAG 900
DB 427 GINGIULYTRIALARGMETVALIGINASPLEUGINASNARGSERHTHREULYSGLIUGIUL 446
QY 901 TCCCTGACACTAACA-----GACTGCTGGGCGAATCTGAGCACCACACTGCAC 948
DB 447 ILELYSGIULIETHTSERSEPHLEUGIULYSILEHTHRSPLEULYSASNGLINLEUARG 466
QY 949 CTCACCAAGATCTGGAATA---GCTCACCTCAGTCAAGAAAGGATACATATGATATA 1005
DB 467 GINGIULASPRGLIULASPHARGIULYSLEUGIULYSGIULYSLYARGHTHRIAGIULYS 486
QY 1006 TTGGGAAAGTTACAGAGAAGAAATGAGAAATGAGAACTGTGTCCAGCATGGAGAGA 1065
DB 487 GIULASNVALMETHTHGLIUEUTHRMETGIULILEASNLYSTRIPARGLEULYTYRASNSP 506
QY 1066 GTACATGACACATGATAGCAAAAGGCTAAGCGACGTGGAT----- 1104
DB 507 LEUTYRGIULYTHRLYSPHTPHGGINGLINGINLEUASPRALPHGGLIULAGIULYSGLIN 526
QY 1105 -----AAGCAGACGCCAGCGCACGCCGACACTG----- 1134
DB 527 ALALEULEUSNGLIHISGLY---ALATHRGINGIUGINLEUASNLYSILEARGASPSER 545
QY 1135 ---GTGACGCTCTCAGCAAGCAGAAC-----CAGCTT 1164
DB 546 TYRIALAGINLEULEUGIYHISGLINASNLEULYSGLINLYSILEYSHISVALIYALYSLEU 565
QY 1165 CTCCTGGAGAGGACAGACCTGTGCGAAGAGGTGACCGGCTCGGACCCAGTTACCACAC 1224
DB 566 LYSASPRGIULASNSERGINLEULYSSERGIUVALISERLYSLEUARGSERGINLEUVALYS 585
QY 1225 ATGCCCAACATCTGAT 1239
DB 586 ARGLYSGIULASNGLIU 590

```

RESULT 2  
US-09-104-324B-4  
Sequence 4, Application US/09104324B  
Patent No. 6232460  
GENERAL INFORMATION:  
APPLICANT: T recti, Ozlem; Sahin, Ugur; Pfeundschnh, Michael  
TITLE OF INVENTION: Methods for Diagnosis and Treating Cancers,  
TITLE OF INVENTION: And Methods for Identifying Pathogenic Markers In A Sample  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fulbright & Jaworski LLP  
STREET: 666 Fifth Avenue  
CITY: New York City  
STATE: New York  
ZIP: 10103  
COMPUTER READABLE FORM:

```

1      MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
2
3      COMPUTER: IBM
4
5      OPERATING SYSTEM: PC-DOS
6
7      SOFTWARE: Wordperfect
8
9      CURRENT APPLICATION DATA:
10
11     APPLICATION NUMBER: US/09/104,324B
12
13     FILING DATE: 25-June-1998
14
15     CLASSIFICATION: 435
16
17     PRIOR APPLICATION DATA:
18
19     APPLICATION NUMBER: 08/892,702
20
21     FILING DATE: 15-July-1997
22
23     ATTORNEY/AGENT INFORMATION:
24
25     NAME: Hanson, No. 623246Oman D.
26
27     REGISTRATION NUMBER: 30,946
28
29     REFERENCE/DOCKET NUMBER: IUD 5491
30
31     TELECOMMUNICATION INFORMATION:
32
33     TELEPHONE: (212) 318-3000
34
35     TELEFAX: (212) 752-5958
36
37     INFORMATION FOR SEQ ID NO: 4:
38
39     SEQUENCE CHARACTERISTICS:
40
41     LENGTH: 976 amino acids
42
43     TYPE: amino acid
44
45     TOPOLOGY: linear
46
47     US-09-104-324B-4

```

Alignment Scores:	
Pred. NO.:	1.91e-13
Score:	232.50
Percent Similarity:	42.60%
Best Local Similarity:	22.08%
Query Match:	8.51%
DB:	4
	Gaps:
US-09-502-945-1 (1-1552) x US-09-104-324B-4 (1-976)	976
	Length:
	Matches:
	Conservative:
	100
	Mismatches:
	147
	Indels:
	113
	Gaps:
	18

US-09-502-945-1 (1-1552) x US-09-104-324B-4 (1-976)

QY	16	GAGACAGCTTAAACTACTTATGAGGAAAGCTGAATTTGAGCAATCTTAAGCTT	75
Db	388	GlutIuLeuLeuAthrInglInglInglAargLeuIuLysAsnGluAspGlnLeuLysIle	407
QY	76	TTGAGGACGACTTAACTGAA-----	96
Db	408	LeuThrMetGluLeuGlnLysLysSerSerGluLeuGluGluMetThrLysLeuThrAsn	427
QY	97	---TATCAGAGAACTTGAGAGATCTTAAAGACGACTTAAAGCATTAAGAATTTCTGTG	153
		:::	
Db	428	AsnLysGluValGluLeuGlnGluLeuLysValLeuGlnGlnLysGluThrLeu	447
QY	154	GCTGCTAATACTTGTAACCGTGTGGTCTTTGTTGAATGTGCTCAGCATGAAGCT	213
Db	448	TyrGluAsn-----LysGlnPheGluLys	455
QY	214	GTT-----CTTCCCAACCCATCACTAATGTTCATATGCAAGCCATCGAAAGCTG	264
		:::	::: ::::
Db	456	IleIaGluGluLeuLysGlyThrGluGlnGluLeu---IleGlyLeuGlnGlnIaIaArg	474
QY	265	GTTTAAAGAAAGAGATGACTTGATGTCTGCAGTATGTTCCGTAAGGACAGCTGGCAGAT	324
			::::
Db	475	GluLysGluValHisAspLeuGluIleGlnLeuThrAlaIleThrIleThrSer	491
QY	325	ACGCAGCAAGAGACAGCAAGTCTTATGAACAGGTGAAACAGTTTGGCAAAATATCTGAG	384
Db	492	-----GluGlnTyrTyrSerLysGluValLysAspLeuLysThrIngluLeuGlu	507
QY	385	GAAGCCAAATTTGAAAAAACCAAGGCTTTATCCAGTGTGACCAAGTTG-----AGG	435
		:::	
Db	508	AsnGlnLysLeuLysAsnThrIngluLeuThrSerHisScyAsnLysLeuSerLeuGluAsn	527
QY	436	AAGGACGTGAGAGAGCGAGCGAGACGACTTGAATAAACCTTGATCTGCAAGAGAAA	495
			:::
Db	528	LysGluLeuThrInglIngluThrSerAspMetCthrLeuGlnLysAsnGlnGlnGlu---	546
QY	496	AGGGCCATTGAGAAAGACATGATGAAATAAGGAAATTAACGAAGAAGAGAGTACATGGGA	555
			:::

Db	547	-----AspIleasnAsnLysLysGlnGlnGluLys-----	557
Oy	556	TCGAAGATGTTGATCTTGCTCAGAAATATTGCCCAACTGAGGCCAGCTGGAAAAAGTT	615
Db	558	-----MetLeuLysGlnIleGluAsnLeuGlnGlnThrGlnThrGlnLeuArg-----	573
Oy	616	ACAAGGAAAAAGATTTCAGCTATTATCAACCTGGAGGAATTCAAAGCCAGCTGCTTCT	675
Db	574	-----AsnGluLeuGlnIleValArgLysGlnLysGln	585
Oy	676	CGGAAATGCAATGTC-----ACAAAGGTGTGGAGAAAAATGGCG	714
		:::	:::
Db	586	LysArgAspGlnValLysCysLysLeuAspLysSerGlnGluAsnLysAsnLeuArg	605
Oy	715	TATAGCTGGAATTAACCAACATGGAGAAAGATGACGCAAAAAAGAGCAGCAGAGATTC	774
Db	606	LysGlnValGluAsnLysAsnLysTyrIleGlnGlnLeuGlnGlnGlnLysAlaLeu	625
Oy	775	AGAGCAAAA-----ACTAACAGGAGATCTTGAATTAAGATCAGGAATATAGAGAA	825
		:::	:::
Db	626	LysLysLysGlyThrAlaGlnSerLysGlnLeuAsnValTyrGlnIleLysValAsnLys	645
Oy	826	TTGAGATATAGACCTGATGTAACCAACAA-----	855
Db	646	LeuGlnLeuGlnLeuGlnSerAlaLysGlnLysPheGlyGlnIleThrAspThrTyrGln	665
Oy	856	-----CACTTGAACAGACGACGACAGAGCA	882
		:::	:::
Db	666	LysGlnIleGlnAspLysLysIleSerGlnGlnLysLeuLeuGlnGlnValGlnLysAla	685
Oy	883	GCCCTGGCCAGAGAGAGCTCGAGACTA--ACAGAACTGCTGGCGCAATCTGAGCAC	939
		:::	:::
Db	686	LysValIleAlaAspGlnAlaValLysLeuGlnLysGlnIleAspLysArgCysGlnHis	705
Oy	940	CAATGCACCTCACCAATGTAATGATGCAACCTGATGCAAGAAAAAGATATCATAT	999
Db	706	LysIle-----AlaGlnMetValAlaLeuMetGlnLysHisLysHisGlnTyr	721
		:::	:::
Oy	1000	GATAAATTGGGAAAGTTACAGAGAGAAATGAGAATGGAGTAATGGAGAACAGTGTCTCCAGAT	1059
Db	722	AspLysIle-----IleGlnGlnArgAspSerGlnLeu-----	732
Oy	1060	GGGAGATACATGACGATGAGAGCAAGGCTATGAGCAGCTGATATAGACAGCACAGGCC	1119
Db	733	---GlyLeuTyrLysSerLysGlnGlnGlnSerSerLeuArgAlaSerLeuIle	751
Oy	1120	ACAGCCAGCAGACGTG--GTGACAGCTCTCTAGCAACAGACAGCTTCTCTGGAGAG	1176
Db	752	GluLeuSerAsnLeuLysAlaGlnLeuLeuSerValLysGlnLeuGlnIleGluArg	771
Oy	1177	CAGAGCTGTGCGAAGAGGTGGACCGGCTGGCAGCCAG	1215
Db	772	-----GlnGlnLysGlnLysLeuLysArgGln	780
		-----	-----
	RESULT 3		
	US-08-328-254-6		
	Sequence 6, Application US/08328254		
	Patent No. 5710022		
	GENERAL INFORMATION:		
	APPLICANT: Zhu, Xueliang		
	APPLICANT: Lee, Wen-Hwa		
	TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein		
	NUMBER OF SEQUENCES: 8		
	CORRESPONDENCE ADDRESS:		
	ADDRESSEE: Campbell and Flores		
	STREET: 4370 La Jolla Village Drive, Suite 700		
	CITY: San Diego		
	STATE: California		
	COUNTRY: USA		
	ZIP: 92122		
	COMPUTER READABLE FORM:		
	MEDIUM TYPE: Floppy disk		
	COMPUTER: IBM PC Compatible		



TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN  
 NUMBER OF SEQUENCES: 60  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: FISH & NEAVE  
 STREET: 1251 AVENUE OF THE AMERICAS  
 CITY: NEW YORK  
 STATE: NEW YORK  
 COUNTRY: U.S.A.  
 ZIP: 10020-1104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Rel. #1.0, ASCII  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US-08/477,831C  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 514  
 ATTORNEY/AGENT INFORMATION:  
 NAME: PIERRI, MARGARET A.  
 REGISTRATION NUMBER: 30,709  
 REFERENCE/DOCKET NUMBER: SIM-10  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-596-9000  
 TELEFAX: 212-596-9090  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 606 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 DESCRIPTION: /desc = "RHAMM I protein"  
 US-08-477-831C-2

Alignment Scores:  
 Pred. No.: 1,74e-13 Length: 606  
 Score: 232.00 Matches: 120  
 Percent Similarity: 41.51% Conservative: 100  
 Best local Similarity: 22.64% Mismatches: 174  
 Query Match: 8.49% Gaps: 136  
 DB: 4 Indels: 25

US-09-502-945-1 (1-1552) x US-08-477-831C-2 (1-606)

QY 4 CTGATGATCCGAGAGCTAAACTTACTATGAGAAAGTGGAATGTGAGAAATCC 63  
 DB 54 LeucysAlaSeraspGlnVal-----GluLysCysLysValAspIleAla 68  
 QY 64 CAATGGAAGTTTGGAGAGACCTAGCTGAATATACAGAACTGTGAGATCTTAAA 123  
 DB 69 GlnLeuGlu-----GluAspLeuLysGlnLysAspArgGluIleLeuSerLeuLys 85  
 QY 124 GAGCACTAAAGCATAAAGATTTCTTCTGCTGCTAATACTTGAACCGTGTGGTGT 183  
 DB 86 GlnSerLeuGlnGluAsn-----IleThrPheSerLysGlnIleGluAsp 100  
 QY 184 CTTTGTTGGAATGT-----GCTCAGCATGAGCTGTCTTCCCAAAACCAT 231  
 DB 101 LeuThrValLysCysGlnLeuLeuGlnThrGluArgAspAsnLeuValSerLysAspArg 120  
 QY 232 -----ACTAATGTCATATGACAGACATC--GAAGACGTGTTAAAGAA 273  
 DB 121 GluArgAlaGlnThrLeuSerAlaGlnMetGlnIleLeuThrGluArgLeuAlaLeuGln 140  
 QY 274 AGAGATGAC-----TTGATGCTGTGCACACTAGTTCCGTAAGAGACGCTTGACAGAT 324  
 DB 141 ArgGlnGluArgGlnLysLeuGlnGlnLysGlnLeuGlnSerGlnSerLeuLeu----- 158  
 QY 325 ACGGAGCAAAAGAGACAGCTTATGACAGAGGTGAAGACAGTTTGCAAAATATCTGAG 384  
 DB 159 GlnGlnGlnLysGlnLeuSerAlaArgLeuGln--GlnGlnLeuCysSerPheGlnGln 177

QY 385 GAACCAATTTGAAAAACC-----AAGGCTTAATCCAGTGT 423  
 DB 178 GluMetThrSerGlnLysAsnValPheLysGlnGluLeuLysLeuAlaLeuAlaGlnLeu 197  
 QY 424 GACCAGTTGAGGAAGAGCTGAGAGCGGAGCGACCTTGAAAAAGAACTTGACATCT 483  
 DB 198 AspAlaValGlnGlnLys---GlnGlnGlnSerGlnLeuValLysGlnLeuGlnGln 216  
 QY 484 CAGCAGAGGAAAGGGCC-----ATTGGAAGAAAGCATGATAAAGGAA 528  
 DB 217 GluArgLysSerThrAlaGlnGlnLeuThrArgLeuAspAsnLeuLeuArgGlnLysGln 236  
 QY 529 ATAAGAAAGAAAGAGAGTACATGAGATCAAGATGTTGATCTTG----- 573  
 DB 237 ValGlnLeuGlnLysHisIleAlaHisIleAlaHisIleAlaGlnIleAlaGlnLys 256  
 QY 574 -----TCTCAGATATTTGCCCACTGGAGGCCCGGAGGAAAGGTTACAAAG 621  
 DB 257 TyrAsnAspThrAlaGlnSerLeuArgAspValThrAlaGlnLeuGlnSerValGlnGln 276  
 QY 622 GAAAAGATTTCAGCTATTATCACTGAGAGGAAATTCAAAGCGCGCTTCGGGAA 681  
 DB 277 LysTyrAsnAspThrAlaGlnSerLeuArgAspValThrAlaGlnLeuGlnSerGlnGln 296  
 QY 682 -----ATGATGTCACAAAGCTGTGTGAGAAATG 711  
 DB 297 GluLysTyrAsnAspThrAlaGlnSerLeuArgAspValThrAlaGlnLeuGlnSerGln 316  
 QY 712 CGCTATCAGCTGAATTAACCAACATGAG--AAGAT-----GAGGCA 753  
 DB 317 GlnGlnLysTyrAsnAspThrAlaGlnSerLeuArgAspValThrAlaGlnLeuGlnSer 336  
 QY 754 GAAAAGGAGCAGACAGACTTACAGCAAAACTAAGCGGATCTT----- 798  
 DB 337 ValGlnGlnLysTyrAsnAspThrAlaGlnSerLeuArgAspValThrAlaGlnLeuGln 356  
 QY 799 -----GAAATTAAGATCAGAAATGAGAAATGAGAAATGAGAAATG 834  
 DB 357 SerTyrLysSerSerThrLeuLysGlnIleGlnLysLeuLysLeuGlnLysLeuThrLeu 376  
 QY 835 GAA-----CTGGATGAAAGCAACACACTTGGAACAGAG----- 870  
 DB 377 GlnGlnLysValAlaMetAlaGlnLysSerValGlnAspValGlnGlnGlnIleLeuThr 396  
 QY 871 -----CAGCAGAAAGGAGCC 885  
 DB 397 AlaGlnSerThrAsnGlnGlnLysThrAlaArgMetValGlnAspLeuGlnAsnArgSerThr 416  
 QY 886 CTGGCCAGAGAGAGTGCCTGAGACTAACA-----GAACCTGCTGGCGAAATCT 933  
 DB 417 LeuLysGlnGlnGlnIleLysGlnIleThrSerSerPheLeuGlnLysIleThrAspLeu 436  
 QY 934 GAGCACAACCTGCACCTCACCAGATCTGAATA--GCTCACTCAGTCAAGAAAAAGG 990  
 DB 437 LysAsnGlnLeuArgGlnGlnAspGlnAspPheArgLysGlnLeuGlnLysGlnLys 456  
 QY 991 TATACATATGATTAATTTGGGAAGTTACAGAGAAAGAAATGAAGATTGGAGAACAGTGT 1050  
 DB 457 ArgThrAlaGlnLysGlnAsnValMetThrGlnLeuThrMetGlnIleAsnLysTrpArg 476  
 QY 1051 GTCCAGCATGGAGAGTACATGAGACATGAGCAAGGCTTAAGCAGCTGAT----- 1104  
 DB 477 LeuLeuTyrAspGlnLeuTyrGlnLysThrLysProPheGlnGlnGlnIleAspAlaPhe 496  
 QY 1105 -----AAGCAGCCAGGCGCAGACGCTGAGCTG-- 1134  
 DB 497 GluAlaGlnLysGlnAlaLeuLeuAsnGlnHisGly--AlaThrGlnGlnGlnLeuAsn 515  
 QY 1135 -----GTCCAGCTCTCAGCAGAGAGAA-- 1158  
 DB 516 LysIleLeuArgAspSerTyrAlaGlnLeuLeuGlnLysHisGlnIleAsnLysGlnLysIleLys 535  
 QY 1159 -----CAGCTTCTCTGGAGAGGAGAGCTGTGCGAAGAGGTGAGACGGCTCGG 1209

```

Db      536 HisValValLysLeuLysAspGluAsnSerGlnLeuLysSerGlnValSerLysLeuArg 555
QY      1210 ACCGAGTTACCAAGCATGCCACATCGAT 1239
Db      556 SerGlnLeuValLysArgLysGlnAsnGlu 565

RESULT 5
US-09-310-187A-1
; Sequence 1, Application US/09310187A
; Patent No. 6358751
; GENERAL INFORMATION:
; APPLICANT: Benichou, Gilles
; APPLICANT: Fedoseyeva, Eugenia
; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac
; FILE OF INVENTION: Graft Rejection
; FILE REFERENCE: UCSF-090
; CURRENT APPLICATION NUMBER: US/09/310,187A
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-310-187A-1

Alignment Scores:
Pred. No.:      3,94e-13      Length:      1939
Score:          230.50        Matches:      109
Percent Similarity: 42.95%    Conservative:  89
Best Local Similarity: 23.64%  Mismatches:   178
Query Match:     8.44%       Indels:       85
                        Gaps:       17
DB:

US-09-502-945-1 (1-1552) x US-09-310-187A-1 (1-1939)
QY      25  AACCTTACTTATGAGCAAAAGTGT----- 48
Db      993  LysLeuThrLysGlnLysLysAlaLeuGlnGlnAlaHisGlnGlnAlaLeuAspLysLeu 1012
QY      49  GAAATTTGAGCAATCCCAATGAGTTTGTGAGACAGACTTACCTGGAATTCAGAGACT 108
Db      1013 GlnValGlnLysLysLysValAsnSerLeuSerLysSerLysValLysLeuGlnGln 1032
QY      109  TGTGAAGATCTTAAAGACAACTA---AAGCATAAAGAAATTTCTGCGCTGCTAATATCT 165
Db      1033 ValAspAspLeuGlnLysSerLeuGlnGlnLysLysValArgMetAspLeuGlnArg 1052
QY      166  TGTAAACCGTGTGGTCTTTTGTGAAATGTGCTCAGCATGAGCTGTTCTTCCCAA 225
Db      1053 AlaLysArgLysLeuGlnLysAspLeuLysLeuThrGln---GlnSerIleMetAspLeu 1071
QY      226 ACCCATACTAATGTCATATGACAGACCATCGAAGACCTGTTAAAGAAAGATGACTTG 285
Db      1072 -----GlnAsnAspLysLeuGlnLysLeuGlnLysLysLysGlnPheAspIle 1089
QY      286 ATGCTCTGACTAGTTCCTGTAAGAGACAGCTTGCGATACCGCAAGAAAGAGCAAGT 345
Db      1090 AsnGlnGlnAsnSerLysIleGlnAspGlnGlnAlaLeuAlaLeuGlnLeuGlnLys 1109
QY      346 GCTTATGAACAGGTGAACAGATTTTGCAAAATATCTGAGAGACCAATTTTGAAGAAAC 405
Db      1110 LeuLysGlnAsnGlnAlaArgIleGlnGlnLysLeuGlnLysLeuGlnLysArgThr 1129
QY      406 AAGCCTTTATCCAGTGTGACCGAGTGAGG-----AAGAGCTGGAGAGCGAG 453
Db      1130 ---AlaArgAlaLysValLysLysLeuArgSerAspLeuSerArgLysLeuGlnGln 1148
QY      454 GCGAGACGAGCTGGA----- 468
Db      1149 SerGlnArgLeuGlnLysAlaGlnLysAlaThrSerValGlnIleGlnMetAsnLysLys 1168

```

```

QY      469 AAGAACTTGATCATCTGACGAAGAAAGGCCATTGAGAAAGACATGATGAAAAAGAA 538
Db      1169 ArgGlnAlaGlnLysLeuGlnLysMetArgAspLeuGlnLysLeuThrLeuGlnHisGln 1188
QY      529 ATAACGAAAGAA-----AGGAGTACATGGGATCAAAAGATGTGATCTTGCTCAGAA 582
Db      1189 AlaThrAlaAlaLeuArgLysLysHisAlaAspSerValAlaGlnLeuGlnGlnGln 1208
QY      583 ATTGCCCACTGGAGGCCCGAGTGAAGAAAGCTTACAAAGGAAAGAAATTTACGATTAAT 642
Db      1209 IleAspAsnLeuGlnArgValLysGlnLysLeuGlnLysLysSerGlnPheLysLeu 1228
QY      643 CAACTGGAGGAAATTCAAAC-----CACTGGCTTCTCGGGAATGATGATGCACA 693
Db      1229 GlnLeuAspAspValThrSerAsnMetGlnGlnIleLeuLysAlaLysAlaAsnLeuGln 1248
QY      694 AAGGTGTGTGAGAATGCGCTTACAGCTGAATATTAACCAACTGGAGAGATGAGCGCA 753
Db      1249 LysValSerArgThrLeuGlnAspGlnAlaAsnGlnLysArgValLysLeuGlnGln 1268
QY      754 GAAAGAGACACAGAGAGTTC-----AGAGCAAAACTAAC----- 789
Db      1269 GlnArgSerLeuAsnAspPheThrThrGlnArgAlaLysLeuGlnThrGlnAsnGln 1288
QY      790 -----AGGATCTTGAATTAAGATCAGCAAAATGAGAAATG----- 828
Db      1289 LeuAlaArgLysLeuGlnGlnLysGlnValAlaLeuIleSerGlnLeuThrArgGlyLysLeu 1308
QY      829 -----AGATAGACATCGATGTAAGCAACCAACTGTGACAGAG-----CAGCAGAG 879
Db      1309 SerTyrThrGlnGlnMetGlnAspLeuLysArgLysLeuGlnGlnGlnLysLysAlaLys 1328
QY      880 GCAGCCCTGGCC----- 900
Db      1329 AsnAlaLeuAlaHisAlaLeuGlnSerAlaArgHisAspCysAspLeuLeuArgGlnGln 1348
QY      901 TGCCTGAGACTAACAGAACTGCTGGCGGAATCTGAGCACCACCACTGCACCTCACCAGATCT 960
Db      1349 TyrGlnGlnLysThrGlnAlaLysAlaGlnLysLeuGlnArgValLeuSerLysAlaAsnSer 1368
QY      961 GAATAGCTCACTCACTCAAGTCAAGAAAGGTATACATATGATTAATGGCAAGTTACAG 1020
Db      1369 GlnValAlaGln-----TrpArgThrLysTyrGlnThrAspAlaIle----- 1382
QY      1021 AGAAGAAATGAGAAATTTGAGAGAACAGTGTGTCACATGCGAGAGATGACATGAGAGATG 1080
Db      1383 GlnArgThrGlnLysLeuGlnGln-----AlaLysLysLysLeu 1395
QY      1081 AAGCAAAAGCTTAAGGACGCTGGATTAAGCACAGCCAGCCACAGCCACAGCAGCTGTGCAG 1140
Db      1396 AlaGlnArgLeuGlnAspAlaGlnLysAlaValGlnAlaValAsnAlaLysCysSerSer 1415
QY      1141 CTCCTCAGACAGACAGAACCGCTTCTCCTGAGAGAGAGAGCGCTGCGAAGAGGTGGAG 1200
Db      1416 LeuGlnLysThrLysHisArgLeuGlnAsnGlnIleLeuAspLeuMetValAspValGln 1435
QY      1201 CCG 1203
Db      1436 Arg 1436

RESULT 6
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DOREMAN, HERRELL AND SKILLMAN

```

STREET: 1601 MARKET STREET, SUITE 720  
 CITY: PHILADELPHIA  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19103-2307  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/353,700  
 FILING DATE: 09-DEC-1994  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: REED, JANET E.  
 REGISTRATION NUMBER: 36,252  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (215) 563-4100  
 TELEFAX: (215) 563-4044  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3248 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: HUMAN  
 US-08-353-700-1

Alignment Scores:  
 Pred. No.: 9.39e-13 Length: 3248  
 Score: 227.50 Matches: 97  
 Percent Similarity: 37.55% Conservative: 93  
 Best Local Similarity: 19.17% Mismatches: 175  
 Query Match: 8.33% Indels: 141  
 Gaps: 15

US-09-502-945-1 (1-1552) x US-08-353-700-1 (1-3248)

QY 1 CTTCTGATGATCCGAGAACCTAAACTTACTTATGAGAAAGTGAATGAGGAA 60  
 Db : : : : : ||||| : : : : :  
 Db 75 ILeYsgIuSerIeugIuLyThrLySgIuLySleSerHISgIuLeuGInValLySgIu 94  
 QY 61 TCCCAATGGAATTTTGGAGAACGACTTACCTGAATATCAGAGAACCTTGAAGATCTT 120  
 Db : : : : : ||||| : : : : :  
 Db 95 SerGInValaSnPhgInGluGlyGInLeuSnsSerGlyLySgIuIleGluLySleu 114  
 QY 121 AAAAGCACTAAAGCAATAAAGATTTCTTGCTGCTAATCTGTGAACCGTGTGT 180  
 Db : : : : : ||||| : : : : :  
 Db 115 GluGInGluLeuLyS----- 119  
 QY 181 GGTCTTTGTTGAATGTGCTCAGCATGAGAGCTGTTCTTCCAAACCATATATGTT 240  
 Db : : : : : ||||| : : : : :  
 Db 120 -----ArgCysLyS----- 122  
 QY 241 CATATGCAACCATGAAAGACTGGTTAAAGAAAGATGACTTGATGCTGCAGTATT 300  
 Db : : : : : ||||| : : : : :  
 Db 123 -----SerGluLeuGluArgSerGInGlnaIaInSerAlaSprVal 137  
 QY 301 TCCGTAAAG-----AGCAGCTTGGCGAGATACCGAG 330  
 Db : : : : : ||||| : : : : :  
 Db 138 SerIeuSnPocYsaSnThrProGInLySlePheThrThrProLeuThrProSerGIn 157  
 QY 331 CAAAGAGAACAGCTGCTTATGACAGAGTGAACAAAGTTTG-----CAATATCTGAG 384  
 Db : : : : : ||||| : : : : :  
 Db 158 TyrItyrSerGlySerLySlyGluAspLeuLySgIuLySlyTyrAsnLySgIuValGluGlu 177  
 QY 385 GAAGCAATTTTGA-----AAAACCAAGGCTTTA 414  
 Db : : : : : ||||| : : : : :

Db 178 ArgLySArgLeuGluValaGluValLySAlaLeuGInaIaLySAlaSerGInThrLeu 197  
 QY 415 ATCCAGTGTGACCACTGAGAGAGAGCTGGAGAG----- 450  
 Db : : : : : ||||| : : : : :  
 Db 198 ProGInaIaThrMetAsnHISArgSprIleAlaArgHISGInaIaSerSerValPhe 217  
 QY 451 -----CAGCGGAGCGACTTGAAMAGACTGCATCTCAGCAAGAAAGGCCATT 504  
 Db : : : : : ||||| : : : : :  
 Db 218 SerItyrGInGInGluLyThrProSerHISLeuSerSerAsnSerGInArgThrProIle 237  
 QY 505 GAGAAAGACATGATGAAAGAAATACGAAGAAAGAGAGTCAATGGAATCAAGATG 564  
 Db : : : : : ||||| : : : : :  
 Db 238 ArgArgAspPheSerAlaSerTyrPheSerGlyGluGluValThrProSerArgSer 257  
 QY 565 TTGATC----- 570  
 Db : : : : : ||||| : : : : :  
 Db 258 ThrLeuGInIleGlyLySArgSprAlaAsnSerSerPheGlyAsnSerSerPro 277  
 QY 571 -----TTGCTTCAGAAATATGCCCCAAGT 594  
 Db : : : : : ||||| : : : : :  
 Db 278 HisLeuLeuAspGInLeuLySAlaGInaSnGInGluLeuAlaGAsnLySleAsnGInLeu 297  
 QY 595 GAGGCCAGGTGAGAAAGCTTACAAAGGAAAGATTCAGCTATTATCAACTGAGGAA 654  
 Db : : : : : ||||| : : : : :  
 Db 298 GluLeuArgLeuGInGlyHISGlySgIuSgIuMetLySgIuValaSnLySlePheGInGlu 317  
 QY 655 ATTCAAGCCCAAGCTGGCTTCGGGAAATGATGTC-----ACAAAGTGTGGA 705  
 Db : : : : : ||||| : : : : :  
 Db 318 LeuGInLeuGInLeuGluLySAlaLySAlaLySValGluLeuIleGluLySgIuLySValLeuAsn 337  
 QY 706 GAAATGCGCTATCAGCTGATGAAATGAAACCAATGAGAGATGAGAGCAAGAAAGAGAC 765  
 Db : : : : : ||||| : : : : :  
 Db 338 LysCysArgAspGluLeuValaArgThrThrAlaGInItyrAspGInaIaSerThrLySlyr 357  
 QY 766 AGAGATTGAGAGCAAAAGCTAACAG-----GATCTGAAATTAAGATCAGGAA 816  
 Db : : : : : ||||| : : : : :  
 Db 358 ThrAlaLeuGInGluGInLySLeuLySLeuThrcGluAspLeuSerCysGInArgGInaSn 377  
 QY 817 ATAGAGAAATGAGAAATGAGCTGAT-----GAAAGCAA 852  
 Db : : : : : ||||| : : : : :  
 Db 378 AlaGluSerAlaArgCysSerLeuGInGInLySleLySgIuLySgIuSgIuPheGIn 397  
 QY 853 CAACACTTGAACAGAGCAGACAGAGCAGCCTGGCCAGAGAGAGTGCCTG----- 906  
 Db : : : : : ||||| : : : : :  
 Db 398 GluGluLeuSerArgGInGInArgSerPheGInThrLeuAspGInGluCysIleGInMet 417  
 QY 907 -----AGACTAACAGAACTGCTGGCGAATCTGAGCAGCAACTGCCTCAGCATCT 960  
 Db : : : : : ||||| : : : : :  
 Db 418 LysAlaArgLeuThrGInGluLeuGInGInaIaLySAsnMetHISAsnValLeuGInaIa 437  
 QY 961 GAAATGCTCAAGCTCAGTCAAGAAAGAAAGTATACATATGATTAATTTGGAAAGTTACAG 1020  
 Db : : : : : ||||| : : : : :  
 Db 438 GluLeuAspLySLeuThrSerValLySgIn-----GluLeuGlu 450  
 QY 1021 AGAAGAAATGAAGATTTGAG-----GAACAGTGTGTCCACATGGG 1062  
 Db : : : : : ||||| : : : : :  
 Db 451 AsnAsnLeuGInGluPheLySgInLySleuCysArgAlaGluGInaIaPheGInaIaSer 470  
 QY 1063 AGACTACATGAGAG-----ATGAAGCAAAAGGCTTAAGCAG 1098  
 Db : : : : : ||||| : : : : :  
 Db 471 GInItyrLySgIuAsnGInLeuAlaArgSerMetGluGluMetLySgIuAsnAsnLeu 490  
 QY 1099 CTGATTAAGACACAGCCAGCCAGCCAGAGCTGTGTCAGCTC-----CTCAGC 1149  
 Db : : : : : ||||| : : : : :  
 Db 491 LeuLySerHISSerGInGInLySAlaArgIuValLySHISLeuGluValaGluLeuLyS 510  
 QY 1150 AAGCAGAACACAGCTTCTCTGAGAGAGAGAGCTGCGGAAGAGTGAACCGGCTCGG 1209  
 Db : : : : : ||||| : : : : :  
 Db 511 AsnItyrSgInCysLeuAsnGInSerGInaSnPheAlaGluGluMetLySAlaLyAsn 530  
 QY 1210 ACCCAGTTACCCAGCAT 1227  
 Db : : : : : ||||| : : : : :  
 Db 531 ThrSerGInGluThrMet 536



```

RESULT 7
PCT-US95-16216-1
: Sequence 1, Application PC/TUS9516216
: GENERAL INFORMATION:
: APPLICANT: Yen, Timothy J.
: APPLICANT: Raltner, Jerome B.
: TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
: TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman
: STREET: 1601 Market Street, Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/16216
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/353,700
: FILING DATE: 09-DEC-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ. ID NO. 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3248 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: PCT-US95-16216-1

Alignment Scores:
Pred. No.: 9, 39e-13 Length: 3248
Score: 227.50 Matches: 97
Percent Similarity: 37.55% Conservative: 93
Best Local Similarity: 19.17% Mismatches: 175
Query Match: 8-33% Indels: 141
DB: 5 Gaps: 15

US-09-502-945-1 (1-1552) x PCT-US95-16216-1 (1-3248)
QY 1 CTTCTGATGATCCGAGAGCTAAACTTACTTATGAGGAAAGTGTGAATTGAGAA 60
   ::::::::::::::::::::
DB 75 TlecygluSerleuInlyThrlysgInlylIleSerHISgluleuInlyVallysgIn 94
QY 61 TCCCAATTGAATTTTGGAGGAAGCACTTAACTGAATATCAGAGAACTGTGAAGATCTT 120
   |||||||:::|||||
DB 95 SerGlnValAsnPhgIngluInlyGlnLeuAsnSerGlyLyslysgInlIleGluLysleu 114
QY 121 AAAGACCACTAAAGCATAAAGAAATTTCTTGCGTGCATTAATCTGTAAACCGTGTGGT 180
   ::::::::::::::::::::
DB 115 GluIngluleuLys- 119
QY 181 GGTCTTTGTTGATGCTCAGCATGAAGCTGTCTTCCCAACCATACTAATGTT 240
   ::::|||
DB 120 -----ArgcysLys----- 122
QY 241 CATATGACAGACATCGAAAGACTGTTAAAGAAAGAGATGACTGTATGTCGACATAGTT 300

```

```

DB 123 -----SerGluLeuGlnArgSerGlnGlnAlaLeuInSerAlaAspVal 137
   |||:::
QY 301 TCCGTAAG-----AGCAGCTTGGCAGATACGCG 330
   |||:::
DB 138 SerLeuAsnProCysAsnThrProGlnLysIlePheThrProLeuThrProSerGln 157
   |||:::
QY 331 CAAGAGAAAGCAAGCTCTTATGAACAGGTGAACCAAGTTTGTG-----CAATATTCAG 384
   ||| |||||:::
DB 158 TyrTyrSerGlySerLysTyrGlnAspLeuLysGlnLysTyrAsnLysGlnValGln 177
   ||| |||||:::
QY 385 GAAGCCAAATTTGAA-----AAAGCCAAAGCTTTA 414
   |||
DB 178 ArgLysArgLeuGlnLysAlaValLysAlaLeuGlnAlaLysLysAlaSerGlnThrLeu 197
   |||
QY 415 ATCCAGTGTGACAGCTTGAGGAGAGCGTCGACAGG----- 450
   |||
DB 198 ProGlnAlaThrMetAsnHisArgAspIleValAlaThrHisGlnAlaSerSerValPhe 217
   |||
QY 451 -----CAGCGGAGCGACTTGAAAAAGAACTTCATCTCAGCAGAGAAAGGCCATT 504
   ||| |||||:::
DB 218 SerTrpGlnGlnLysThrProSerHisLeuSerSerAsnSerGlnArgThrProIle 237
   ||| |||||:::
QY 505 GAGAAAGACATGATGAAAAAGCAATACCAAGAAAGAGAGATCATGGATCAAGAG 564
   ::::|||
DB 238 ArgArgAspPheSerAlaSerTyrPheSerGlyLysLeuGlnValThrProSerArgSer 257
   |||
QY 565 TTGATC----- 570
   |||
DB 258 ThrLeuGlnIleGlyLysArgAspAlaAsnSerPhePheGlyLysnSerSerPro 277
   |||
QY 571 -----TTGCTTCAGATATTTGCCCACTG 594
   |||
DB 278 HisLeuLeuAspGlnLeuLysAlaGlnAsnGlnGlnLeuArgAsnLysIleAsnGlnLeu 297
   |||
QY 595 GAGGCCAGGTGAAAAAGTTACAAAGAAAGATTTACACTTATCACTGAGAGAA 654
   ||| |||||:::
DB 298 GluLeuArgLeuGlnLysHisGlnLysGlnMetLysGlyGlnValAsnLysPheGlnGln 317
   |||
QY 655 ATCAAGCCAGCTGCTCTCTCGGAAATGAGATG-----ACAAAGCTGTGGA 705
   |||
DB 318 LeuGlnLeuGlnLeuGlnLysAlaLysValGlnLeuIleGlnLysGlnLysValIleAsn 337
   |||
QY 706 GAAATCGCTATCAGCTGATATAAACCAACATGAGAGAGATGAGGACAGAAAGAGCAC 765
   ||| |||||:::
DB 338 LysCysArgAspGlnLeuValArgThrThrAlaGlnLysArgAspAlaSerThrLysTyr 357
   |||
QY 766 AGAGAGTTAGAGCAAAACTAACAGG-----GATCTTGAATTTAAAGTCAGGA 816
   |||
DB 358 ThrAlaLeuGlnLysLysLeuLysLysLeuThrGlnAspLeuSerCysGlnArgGlnAsn 377
   |||
QY 817 ATAGAGAAATTTGAGATTAAGACTGAT-----GAAACGAA 852
   |||
DB 378 AlaGlnSerAlaArgCysSerLeuGlnGlnLysIleLysGlnLysGlnLysGlnPheGln 397
   |||
QY 853 CAACACTTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 906
   ||| |||||:::
DB 398 GluGlnLeuSerArgGlnGlnArgSerPheGlnThrLeuAspGlnLysGlnIleGlnMet 417
   |||
QY 907 -----AGACTACAGAACTGCTGGCGCAATCTGAGCACCAGCTCAGCAGATCT 960
   ||| |||||:::
DB 418 LysAlaArgLeuThrGlnLysLeuGlnGlnAlaLysAsnMetHisAsnValLeuGlnAla 437
   |||
QY 961 GAAATAGTCACAGCTCAGTCAGAAAGAAAGGATATACATATGATTAATTTGGAAAGCTTAC 1020
   |||
DB 438 GluLeuAspLysLeuThrSerValLysGln-----GlnLeuGln 450
   |||
QY 1021 AGAAGAAATGAAGAAATTTGAG-----GAACAGTGTGTCAGCATGAG 1062
   |||||
DB 451 AsnAsnLeuGlnGlnLysPheLysGlnLysLeuCysArgAlaGlnGlnAlaPheGlnAlaSer 470
   |||
QY 1063 AGAGTACATGAGACG-----ATGAGCAAAAGGCTAAGGACG 1098
   |||||

```

Dh 471 GlnIleYsgluAsnGlnIleuAArgSerMetGluMetLysGluuAsnAsnIleu 490  
Qy 1099 CTGATTAAGCACACCCAGCCAGCCAGCCAGCTGCTGCAGCTC-----CTCAGC 1149  
Dh 491 LeuYsserHisSerGlnIleuLysAlaArgGluValCysHisLeuGluAlaGluIleuLys 510  
Qy 1150 AAGCAGACACAGCTCTCTGAGAGGAGGAGCCCTGCTGAGAGCTGAGACGGCTGGCG 1209  
Dh 511 AsnIleYsglnCysLeuAsnGlnSerIleAsnPhenAlaGluIleuMetLysAlaLysAsn 530  
Qy 1210 ACCAGATTACCCAGCATG 1227  
Dh 531 ThrSerGlnIleuThreMet 536  
RESULT 8  
US-09-572-191-2  
Sequence 2, Application US/09572191  
Patent No. 6355466  
GENERAL INFORMATION:  
APPLICANT: Beraud, Christophe  
APPLICANT: Sakowicz, Roman  
TITLE OF INVENTION: No. 6355466el motor proteins and methods for  
FILE REFERENCE: 1017  
CURRENT APPLICATION NUMBER: US/09/572,191  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 1388  
TYPE: PRT  
ORGANISM: Human  
US-09-572-191-2  
Alignment Scores:  
Pred. No.: 1,39e-12 Length: 1388  
Score: 224.00 Matches: 112  
Percent Similarity: 38.36% Conservatrive: 94  
Best Local Similarity: 20.86% Mismatches: 183  
Query Match: 8.20% Indels: 148  
Gaps: 17  
US-09-502-945-1 (1-1552) x US-09-572-191-2 (1-1388)  
Qy 1 CTTCGTGATGCATCCGAACCTAAACTTACTTATGAGAAAGTGAATTCAGAGAA 60  
Dh 856 LeuLeuGlnSerLysAlaCysLeuGlnAspSerTyraSpasnLeuGlnIleuMetLys 875  
Qy 61 TCCCAATTGAAGTTTGTGAGACGACTTAGCTGAATTCAGAGAACTTGTGAAGATCTT 120  
Dh 876 PheGluIleAspGlnLeuSerIleAsnLeuGlnAsnPhenLysGluAsnGluThreIleu 895  
Qy 121 AAAGACAGCTAAAGCATAAAGAAATTTCTGCTGCTGAATTAATCTTGAACCGTTGGT 180  
Dh 896 LysSerAspLeuAsnAsnLeuMetGluLeuGlnAlaGluLys---GluAArgAsnAsn 914  
Qy 181 GGTCTGTTGTTG----- 192  
Dh 915 LysLeuSerLeuGlnPheGluGlnAspLysGluAsnSerSerLysGluIleuLysVal 934  
Qy 193 -----AAATGCTCAGCATGAGCT 213  
Dh 935 LeuGluAlaValArgGlnGlnLysGlnIleuThralAlaLysCysGlnGlnIleuMetAla 954  
Qy 214 GTTCTTCCCAAAACCATTAATGTT-----CATATGCAGACCATGGAAGACTGGTT 267  
Dh 955 LysValGlnLysLeuGlnIleuSerLeuLeuAlaThrGluLysValIleSerSerLeuGln 974  
Qy 268 AAAGAAAGAGAT-----GACTGATGCTCGCACTACTTCCGTA 306  
Dh 975 LysSerArgAspSerAspLysLysValAlaAspLeuMetAsnGlnIleGlnIleu 994

Qy 307 AGGACAGCTTGGCAGATACCGACGAAAGACGCAAGCTGTTATGACAGTGAACAA 366  
Dh 995 ArgSerValCysGln-----LysThrGluThrIleAspThrLeuLysGln 1010  
Qy 367 GTTTTGCAAATATCTGAGAAAGCCATTTTGAAGAAACCAAGCTTTATCCAGTGTAC 426  
Dh 1011 GluLeu-----LysAspIleAsnCysLysTyraSerAlaIleuValAspArgGlu 1027  
Qy 427 CAGTTGAGGAAGAGACTGTGAGAGCGAG-----GCCGAGCGACTT 465  
Dh 1028 GluSerArgValIleuIleLysLysGlnIleuValAspIleLeuAspLeuLysGluThrLeu 1047  
Qy 466 GAAAGAAAGACTTGCATCTCAGCAAGAAAGGCCCATGTGAAGACATGATGAAAGAG 525  
Dh 1048 ArgLeuArgIleLeuSerGlnAsp-----IleGluAArgAspMetLeuCysGln 1063  
Qy 526 GAAATTAACGAAAGAAAGGAGCTACATGGATCAAGATGTTGATCTTCCAGATATT 585  
Dh 1064 AspLeuAlaHisAlaThrGluGlnIleuAsnMetLeuThrGluAlaSerLysLysHisSer 1083  
Qy 586 GCCCACTGGAGAGCCCACTGAGAAAGCTTACAAAGAAAGATTTCAGCTATTATCA 645  
Dh 1084 GlyLeuLeuGlnSerIleAlaGlnIleuGluThrLysGluAla-----Leu 1099  
Qy 646 CTGGAGAAATTCAAAGCCAGCTGCTCTGCGGAAATGATGTCAAAAGGTGTGGA 705  
Dh 1100 IleGlnGlnIleuGlnHisLysLeuAsnGlnLysGlnIleuValGlnGlnLysAsn 1119  
Qy 706 GAAATGCGCTATCAGCTGAATTAACAACATGGAAGATGAGCA----- 753  
Dh 1120 GluTyraPheLysMetArgGlnIleuGlnHisValMetAspSerAlaIleGluAspPro 1139  
Qy 753 ----- 753  
Dh 1140 GlnSerProLysThrProThrHisPheGlnThrHisIleuAlaLysLeuLeuGluThrGln 1159  
Qy 754 GAAAGAGACACAGAGCTTCAGAGCA-----AAA 783  
Dh 1160 GlnGlnGlnIleGlnAspGlyArgAlaSerLysThrSerLeuGlnHisLeuValThrLys 1179  
Qy 784 ACTAAGAGGATCTTGAATTAAGATCAGGAATA----- 819  
Dh 1180 LeuAsnGlnAspArgGlnValLysAsnAlaGluIleLeuArgMetLysGluGlnIleuArg 1199  
Qy 820 -----GAGAAATTCAGAAATGAA----- 837  
Dh 1200 GluMetGluAsnLeuArgLeuGlnSerGlnGlnIleuIleGluLysAsnThrLeuGln 1219  
Qy 838 -----CTGATGGAAGAGCA----- 852  
Dh 1220 GlyGlnLeuAspAspIleLysArgGlnLysGluAsnSerAspGlnAsnHisProAspAsn 1239  
Qy 853 CAACACTTGAACAGGACAGCAG-----AAGCAGCCCTGGCCAGA----- 894  
Dh 1240 GlnGlnIleuLysAsnGlnGlnGlnIleuSerIleLysGluArgLeuAlaLysSerLysIle 1259  
Qy 895 ---GAGGAGTGGCTGAGACTAACAAGACTGCTGGGCGAATGCGACCAACTGCACTGC 951  
Dh 1260 ValGlnIleuMetLeuLysMetLysAlaAspLeuGlnGluValGlnSerAlaLeuTyraAsn 1279  
Qy 952 ACCAGATCTGAATAGCTCACTCAGTCAAGAAAAAGTATCATATGATGATTAATGGGA 1011  
Dh 1280 LysGlnMetGlnCysLeuArgMetThrAspGlnValGluArgThrGlnThrLeuGlnSer 1299  
Qy 1012 AAGTTACAGAGCAAAATGAAGATTTGAGAGACAGTGTGCCAGCATGGAGAGACTACAT 1071  
Dh 1300 LysAlaPheGlnGlnLysGlnGlnIleuArgSerLysLeuGlnIleuMetLysGlnIleuArg 1319  
Qy 1072 GAGACGATGAAGCAAAAGCTGAAGCAGCTGATGAACACACACAGGCCAGCCAGCAG 1131  
Dh 1320 GluArgThrSerGlnGlnIleuMetLeuArgLysGlnValGluCysLeuAlaGluIleu 1339  
Qy 1132 CTGCTGACGCTCTCAGCAGCAGAAC-----CAGCTT 1164

```

Db      1340  AsnD1yGAGAGAGAGAGAGAGCCCTGTGTGGAGAGAGGTGAGACCGGCTGCGAGCCAG 1215
               ::|||:::  |||||
OY      1165  CTCCTGGAGAGAGAGAGAGAGCCCTGTGTGGAGAGAGGTGAGACCGGCTGCGAGCCAG 1215
               |||  |||:::|||||  :::::|||||  ::
Db      1360  LysLysGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1376
               :::::
RESULT 9
US-09-723-262-2
: Sequence 2, Application US/09723262
: Patent No. 6379912
: GENERAL INFORMATION:
: APPLICANT: Beraud, Christophe
: APPLICANT: Sakowicz, Roman
: APPLICANT: Wood, Kenneth
: TITLE OF INVENTION: No. 6379912el motor proteins and methods for
: TITLE OF INVENTION: Their use
: FILE REFERENCE: 1017
: CURRENT APPLICATION NUMBER: US/09/723,262
: CURRENT FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: US 09/572,191
: PRIOR FILING DATE: 2000-05-17
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: FastSeq for windows Version 4.0
: SEQ ID NO 2
: LENGTH: 1388
: TYPE: PRT
: ORGANISM: Human
: US-09-723-262-2

```

Alignment Scores:	
Pred. No.:	1,39e-12
Score:	224.00
Percent Similarity:	38.36%
Best Local Similarity:	20.86%
Query Match:	8.20%
DB:	4
	Gaps:
	Indels:
	Mismatches:
	Conservative:
	Matches:
	Length:
US-09-502-945-1 (1-1552) × US-09-723-262-2 (1-1388)	1388

OY	1	CTTTCGATGCATCCAGCAAGCTAAACCTTACTATGAGCAAAAAGTGAATATGAGAA	60
Db	856	LeuLeuGluSerLysAlaCysLeuGlnAspSerTyrAspAsnLeuGlnGluLeuMetLys	875
OY	61	TCCCAATGTAAGTTTGGGAGGAGCAAGCACTTACGTGAATATCAGAGACACTTGGAAGACTT	120
Db	876	PheGluIleAspGlnLeuSerLysArgAsnLeuGlnAsnPheLysLysGluAsnGluThr	895
OY	121	AAAGAGCAACTTAAGCATAAAGATTTCTTCGGCGCTAAATACCTGTAAACCGTTGCT	180
Db	896	LysSerLysPheLysAsnAsnLeuMetGluLeuLeuGluAlaGluLys---GluArgAsnAsn	914
OY	181	GGTCTTGTGTTG-----	192
Db	915	LysLeuSerLeuGlnPheGluGlnAspLysGluAsnSerSerLysGluIleLeuLysVal	934
OY	193	-----AAATGGTCGACGATGAACCT	213
Db	935	LeuGluAlaValArgGlnGluLysGlnLysGlnThrAlaLysCysGluGlnGlnMetAla	954
OY	214	GTTCTTTCCCAACCCATCACTATGTT-----CATATGCACACCATCGAAGAAGCTGTT	267
Db	955	LysValAlaGlnLysLeuGlnGluSerLeuLeuAlaThrGluLysValIleLeuSerLeuGlu	974
OY	268	AAAGAAAGAGAT-----GACTTGATGCTCTGCACACTAGTTTCCGTA	306
Db	975	LysSerTrpArgAspSerAspLysLysValValAlaAspLeuMetAsnGlnIleGlnGluLeu	994
OY	307	AGAGACAGCTGTGCACATACGACAGCAAAAGACAGCAAGGCTTAAAGACAGGTGAACAA	366
Db	995	ArgSerSerValLysGlu-----LysThrGlnThrIleAspThrLeuLysGln	1010
OY	367	GTTTTCGCAATATCTGAGAGAGCAATTTTGAACAAAACCAAGGCTTATATCCAGCTGTAC	426

Db	1011	gluleu-----lysasp11leasnlyslytynasnserraleuvalaspargglu	1027
Qy	427	CAAGTTGAGGAAGAGCTGGAGAGCGAG-----GGCGAGCGACTT	465
Db	1028	gluseratrgvalleu11elyslsglgluvalaspl1leuasnpleuysgluthrleu	1047
Qy	466	GAAMAAACAATGCTGCTCAGCAAGAGAAAGGCCATGTAGAAAGACATGTATGAAG	525
Db	1048	argleuarg11leuusergluasn-----11egluargspmetleucysglu	1063
Qy	526	GAATTAACGAAGAAGAGGATCATGGATCGATCAAGATGTATGCTTCAGAAATAT	585
Db	1064	aspleu11anhsalatphrglglu11leuasnmetleuthrglu11asertlys11hsiser	1083
Qy	566	GCCCACTGGAGGCCCGAGTGGAAAAGTTACAAAGAAAGATTTCAAGTATTAATCA	645
Db	1084	glyleu11leu11nsert11ag11ng11glu11leuthrgly11glu11ala-----leu	1099
Qy	646	CTGGAGCAAAATTCAAAGCCAGCTGGCTTCGGGAAATGGATGTCAAAAGGTGTGTGA	705
Db	1100	11eg11ng11leu11ng11nhs11lys11leuasn11glu11nhs11glu11nhs11lysasn	1119
Qy	706	GAATTCGGCTATCGCTGATTAACCAACATCGAGACAGATGAGCA-----	753
Db	1120	glutyrasnpleuysmet11ag11leu11glu11nhs11valmet11aspserr11ala11aspr11	1139
Qy	753	-----	753
Db	1140	gl11nsrpro11ys11thr11pro11nhs11phre11ng11nhs11leu11al11y11leu11eug11uthrg11	1159
Qy	754	GAAGAAGCACAGACAGAGTTCAAGCA-----AAA	783
Db	1160	gl11ng11ng11leu11asparg11y11asert11ys11thr11ser11leu11nhs11leu11val11thr11ys	1179
Qy	784	ACTTAACGGGGCTGTGAATTAAGATTCAGGAAT-----	819
Db	1180	leu11asn11glu11asparg11val11lys11asn11ala11glu11leu11arg11met11ys11glu11leu11arg	1199
Qy	820	-----GAGAAATTGGAGATATGAA-----	837
Db	1200	glu11met11glu11nsleu11arg11leu11g11user11ng11nleu11leu11glu11nhs11ntr11preu11eug11n	1219
Qy	838	-----CTGATGAACAACA-----	852
Db	1220	gl11yn11leu11nhs11aspr111ely11asrg11n11lys11glu11asn11ser11aspserr11nhs11pro11aspr11n	1239
Qy	853	CAACACTTGGACAGGAGCGAGAG-----AAGCGACGCTGGCCGAGA-----	894
Db	1240	gl11n11leu11nhs11asn11ng11ng11ng11user111ely11sg11u11arg11leu11al11by11ser11ys11le	1259
Qy	895	---GAGGATGCGCTGAGACTACACAGAAAGACTGCGGGAGCAATTCAGACCAACTGCACCTC	951
Db	1260	val11glu11leu11met11eul11ys11met11ys11ala11aspleu11ng11u11val11g11nser11al11eul11y11asn	1279
Qy	952	ACCAGATCTGAATTAAGCTCAACTCACTCAAGAAAAAGGTATACATATGATTAATTGGCA	1011
Db	1280	lys11glu11met11nhs11cyl11user11arg11met11tr11asrg11u11val11glu11arg11thr11g11ntr11leu11g11user	1299
Qy	1012	AAGTTACAGGAAGAATTAAGAATTGGAGGAACAGTGTGTCCACAGTGGAGAGTAAAT	1071
Db	1300	lys11al11ar11he11ng11ncl11y11sg11ncl11leu11arg11ser11lys11leu11glu11glu11met11y11rg11u11arg	1319
Qy	1072	GAGACGATGAAAGCAAGAGCTTAAGCGAGCTGGATTAAGACACAGCGAGCGACAGCCAGCA	1131
Db	1320	glu11arg11thr11ser11ng11ncl11met11eul11met11eul11arg11nval11glu11cyl11eul11al11glu11u	1339
Qy	1132	CTGGTCAGCTCTCTCAGCAAGCAAG-----CAAGTT	1164
Db	1340	asn11ncl11y11leu11val11glu11nsg11nhs11nleu11nhs11ng11nhs11leg11ntr11val11al11arg11n	1359
Qy	1165	CTCTGGAGAGGACAGACCTGTCTGGAGAGGAGTGGACCGGCTGTGGAGCCAG	1215

Db	1140	glnserfroulsthrrprogronshrspreghlnthrhlsleualylsleuleuicglwthrgln	1159
Qy	754	GAAGAAAGGCACAGAGAGGTCACAGAGA-----AAA	783
		:::	
Db	1160	gluinegluilegluabspgylatrgdallaSerlysthrSerleuglnhlsleuValthrllys	1179
Qy	784	ACTAACAGGAGATCTTGAATTAAGAATCAGAGAAATP-----	819
		:::     :::	
Db	1180	leuasnclupsratpglnValyllysasnalaIguileleuargMetlysluInleuarg	1199
Qy	820	-----GAGAAATTTGAGCAATPAGAA-----	837
		:::     :::	
Db	1200	gluMetlupasnleuargleugluserGlnGlnleuIlegluLysasnTrpleuLeuGln	1219
Qy	838	-----CTGGATGAAGACAAA-----	852
		:::	
Db	1220	gluInleupsratprrlIelylsatrglnLysgluasnserAspGlnasnhsIproAspasn	1239
Qy	853	CAACACTTTGGACAGACGACGACAG-----AACGCCGCCCTGGCCAGA-----	894
		:::    :::     :::	
Db	1240	glnInleuLysasnnglnnglnGlnuserIlelysluIleargleualylsSerlyslle	1259
Qy	895	---GAGAGAGTCCGTGAGACTAACAGCAACTGCTGGCGGAATCTGAGCACCACCACTGCACCTC	951
		:::     :::        :::        :::	
Db	1260	ValIguInleuMetleuLysMetlyslalaAspserGlnGlnIleuSerAlaleuThylasn	1279
Qy	952	ACCGAGATCTCAATTAAGCTCACTCACTCAACGAAAAAGTATACATATGATTAATTTGGA	1011
		:::     :::	
Db	1280	lysSluMetlucylsleuargMetThrAspGlnuValgluArgThrGlnThrleuGlnser	1299
Qy	1012	AAGTTACAGAGAAAGAAATGAAGAAATGGAGAAACAGTGTCTCAGCATGGGAGATACAT	1071
		:::    :::     :::        :::	
Db	1300	lysAlaIlePhgGlnGlnLysGlnGlnleuArgSerLysleuGlnGlnMetLysGlnGlnArg	1319
Qy	1072	GAGACAGATGAAGCAAGAGGCTTAAGGACGCTGGATTAAGACACGCCACGACGCCACAG	1131
		:::            :::        :::	
Db	1320	gluArgThrSerGlnMetleuMetleuArgLysGlnValGlnCysleuAlaGlnlu	1339
Qy	1132	CTGGTCAGACTCTCTCAGACAGCAGAC-----CACGTT	1164
		:::     :::	
Db	1340	AsnclupLysleuValgluYlnsGlnAsnleuYlnsGlnlylIleGlnTyrValValargleu	1359
Qy	1165	CTCTCGAGAGGACAGAGACCTGTGTGGAGAGAGTGACACGGGCTGGAGCCAG	1215
		:::         :::     :::	



[illegible]









